



PROCESAMIENTO DE IMÁGENES BIOMÉDICAS (16.06)

CARRERA DE BIOINGENIERÍA
INSTITUTO TECNOLÓGICO DE BUENOS AIRES
2021

TP Grupal: Análisis de Lunares

Profesor: Roberto Sebastián TOMÁS
J.T.P: Guillermina GRIFFA
ATP: Iván DUJAUT

Alumnos: Bautista DIPAOLA
Maximiliano LEIBOVICH
Nicolás POSTEGUILLO
Alejo ROMANO

Fecha de Entrega: 23 / 11 / 2021

Índice

| | |
|---|-----------|
| 1. Introducción | 3 |
| 2. Desarrollo | 3 |
| 2.1. Elección de imágenes | 3 |
| 2.2. Selección de características | 3 |
| 2.3. Procesamiento de las imágenes y parámetros básicos | 4 |
| 2.3.1. Recorte | 4 |
| 2.3.2. Segmentación y su mejoramiento | 5 |
| 2.3.3. Parámetros relevantes | 5 |
| 2.4. Algoritmo para cada característica | 6 |
| 2.4.1. Asimetría (A) | 6 |
| 2.4.2. Bordes (B) | 7 |
| 2.4.3. Color (C) | 7 |
| 2.4.4. Diámetro (D) | 8 |
| 3. Resultados | 9 |
| 4. Conclusiones, limitaciones, y futuras mejoras | 9 |
| 5. Anexo: Capturas de la GUI | 11 |
| 6. Referencias | 12 |

1. Introducción

Los lunares son cúmulos de células de la piel, y se producen cuando las células productoras de pigmentos (melanocitos) crecen en grupo. Estos son cuerpos benignos muy comunes y que pueden ir desapareciendo con la edad. Suelen tener características morfológicas similares, en las que destacan un color rosado, moreno o marrón, forma redonda u ovalada, pueden ser planos o presentar una leve elevación, y no superan el tamaño de la goma de un lápiz. Sin embargo, pueden existir algunos lunares que presentan aspectos distintos a los mencionados que tienen mayor potencial de volverse cancerígenos, convirtiéndose en melanomas. [1]

Un melanoma es un tipo de cáncer de piel, es un crecimiento descontrolado de los cúmulos de melanocitos. La mayoría de los melanomas continúan produciendo melanina, por lo que los tumores adquieren diferencias de color, pudiendo llegar a tener distintas gamas de rojo, gris, azul e incluso blanco. Se pueden desarrollar en cualquier parte de la piel, aunque existe una mayor frecuencia de aparición de melanomas en el pecho y espalda en los hombres, y en las piernas de las mujeres. [2]

Si bien los melanomas son mucho menos frecuentes que otros tipos de cáncer de piel, éstos tienen alta peligrosidad ya que pueden propagarse con facilidad a otras áreas del cuerpo si no se descubren y tratan a tiempo [2]. En términos generales, 1 de cada 5 personas en Estados Unidos desarrolla algún tipo de cáncer de piel a lo largo de su vida. Entre los casos de melanoma, si es detectado y tratado a tiempo entonces presenta una tasa de supervivencia del 98 %; mientras que con una detección tardía dicha tasa de supervivencia cae al 15 % [3].

El objetivo de este trabajo fue armar una interfaz gráfica que permita realizar un análisis de las distintas características morfológicas de los lunares, a partir de lo cual poder calcular un *score* y con esto clasificarlos como ‘normales’ o ‘sospechosos’.

2. Desarrollo

2.1. Elección de imágenes

Las imágenes fueron extraídas de la base de datos PH2 [5], que es una gran colección de imágenes de melanomas y de lunares benignos, y es de acceso abierto. Fue creada con propósitos de investigación y de benchmarking para facilitar estudios comparativos y la elaboración de algoritmos de clasificación y segmentación.

Las imágenes se obtuvieron por dermatoscopia, que es una técnica no invasiva en la que se examinan las lesiones dermatológicas mediante un instrumento óptico. Permite ver incluso por debajo de la superficie cutánea, y además su imagen resultante ya está amplificada. Todas se obtuvieron bajo las mismas condiciones, en el Servicio de Dermatología del Hospital Pedro Hispano (Matosinhos, Portugal) y utilizando el sistema *Tuebinger Mole Analyzer* con un aumento de 20x. Son imágenes en color RGB de 8 bits con una resolución de 768x560 píxeles.

Esta base de datos de imágenes contiene un total de 200 imágenes dermatoscópicas de lesiones melanocíticas, incluidos 80 lunares comunes, 80 lunares atípicos y 40 melanomas. La base de datos también incluye un archivo Excel donde se encuentran las anotaciones manuales hechas por los médicos/investigadores para cada imagen, la segmentación de la lesión que realizaron, el diagnóstico clínico e histológico, y la evaluación de varios criterios dermatoscópicos. Esto nos permite evaluar nuestro algoritmo contra estos resultados de referencia, lo cual es importante a la hora de establecer los umbrales para el análisis de cada característica.

2.2. Selección de características

Como sustento teórico para la clasificación de los lunares se optó por utilizar la **Regla ABCD**. Es una técnica popular reconocida por la AAD (American Academy of Dermatology), que se usa en la práctica clínica para clasificar a los lunares en ‘benignos’ o en ‘sospechosos/potenciales melanomas’.

Para aplicar esta regla primero hay que obtener ciertos parámetros o índices que caracterizan al lunar, cada uno representado por una letra y pudiendo tomar un cierto rango de valores discretos. Luego, se ponderan estos valores utilizando la siguiente ecuación:

$$TDS = 1,3 A + 0,1 B + 0,5 C + 0,5 D \quad (1)$$

donde TDS es el puntaje dermatológico total, por sus siglas en inglés. Si este valor es mayor o igual a 5.65, se lo considera sospechoso; mientras que valores inferiores se consideran como lunar normal o sano.

Si bien el rango de valores posibles para cada índice/letra está claramente establecido, la definición de qué representa cada una es sujeto de debate y fluctúa levemente entre las diversas investigaciones (principalmente en lo que respecta al índice 'D'). No obstante, el consenso mayoritario para su definición es el siguiente:

- **Asimetría (A):** Se corresponde con el número de ejes asimétricos que presenta el lunar. Su rango de valores es 0-2, siendo 0 un lunar simétrico, 1 un lunar con asimetría a lo largo de solo uno de sus ejes, y 2 un lunar con asimetría en ambos.
- **Bordes (B):** Se corresponde con el número de segmentos del borde que sean considerados como irregulares. Primero se segmenta la imagen del lunar en 8 partes iguales, y luego se analiza cada una por separado. Aquellas consideradas como irregulares suman 1 al valor de B, mientras que las regulares 0. Por ende, el rango de valores para B es 0-8.
- **Color (C):** Se corresponde con el número de colores que presenta el lunar. Para todo lunar, se considera que sus posibles colores son: Negro, Blanco, Rojo, Marrón Claro, Marrón Oscuro y Gris Azulado. Entonces, el rango de valores para C es 1-6, siendo 1 un lunar que presenta un único color y 6 un lunar que presenta todos los colores.
- **Diámetro (D):** Se corresponde con la medida del diámetro del lunar. Según el valor del diámetro, se calcula un parámetro auxiliar M, y en base a este último se define un rango de 1-5 para el valor de D.

Como ultimo comentario, es interesante mencionar que existen algunas extensiones de la regla ABCD, como por ejemplo la regla ABCDE. La E representa 'Evolución en el tiempo', y es algo importante en la practica clínica. Pero excede a la complejidad buscada para este trabajo, además de que para poder implementarse se necesita una base de datos que tenga fotos del mismo lunar de un mismo paciente a lo largo del tiempo.

2.3. Procesamiento de las imágenes y parámetros básicos

Antes de poder calcular los índices requeridos para la regla ABCD hay que realizar algunos procesamiento previos sobre las imágenes, para que las mismas sean adecuadas. También calcular ciertos parámetros básicos y generales a utilizarse luego.

2.3.1. Recorte

Se realiza un recorte de la imagen original, de manera que se minimice la presencia de piel y que el lunar ocupe la mayor parte del cuadro. Además, también es necesario que en este paso quede fuera todo tipo de 'margen' oscuro en las esquinas de la imagen, propio de la obtención mediante dermatoscopia (ver figura 1). El recorte es realizado "manualmente", es decir, que la interfaz gráfica permite al usuario ingresar la coordenadas que desea incluir.

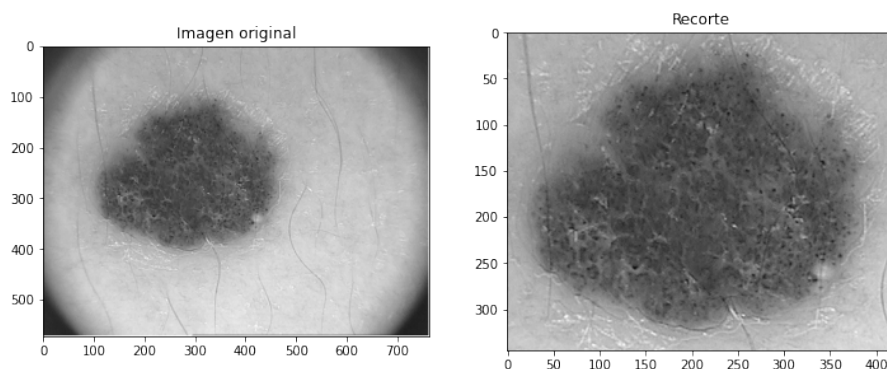


Figura 1: Imagen original en escala de grises (izq.), y su recorte (der.)

2.3.2. Segmentación y su mejoramiento

La segmentación del lunar para ‘separarlo’ del fondo se hace mediante umbralización. El valor de dicho umbral fue hallado mediante el método de Otsu.

Una vez realizada la segmentación, esta puede mejorarse mediante operaciones morfológicas (dilatación, erosión, cierre y apertura) para, por ejemplo, rellenar huecos que hayan quedado. Esto va a depender de cada caso particular, así que es algo que el usuario ejecutaría a través de la interfaz según considere necesario.

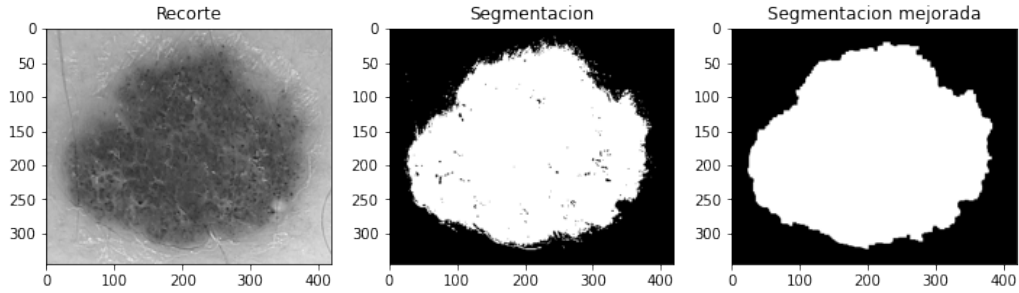


Figura 2: Recorte de la imagen original (izq.), su segmentación mediante umbralización y el método de Otsu (centro), y el resultado del mejoramiento mediante operaciones morfológicas (der.). En este caso, una operación de cierre seguida de una apertura.

2.3.3. Parámetros relevantes

Hay que obtener ciertos parámetros y métricas básicas que luego son utilizadas en el cálculo de los 4 índices.

- **Área:** El número total de píxeles blancos en la imagen segmentada y mejorada (ver figura 2).
- **Perímetro:** Cantidad de píxeles que se encuentran en el borde (de un píxel de grosor) del lunar.

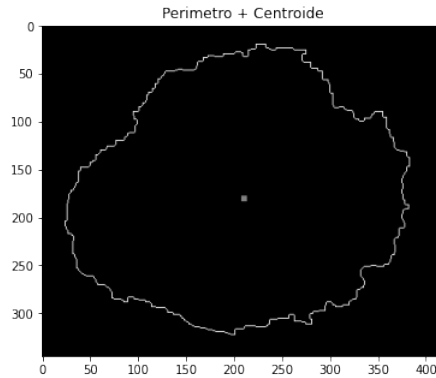


Figura 3: Perímetro y centroide del lunar

- **Centroide:** También llamado centro de masa, sus coordenadas (x,y) se definen como

$$(x_c, y_c) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right)$$

Figura 4: Formula para calcular las coordenadas (x,y) del centroide.

- **Mayor distancia diagonal:** Se define como la recta más larga que pasa entre 2 puntos del perímetro.

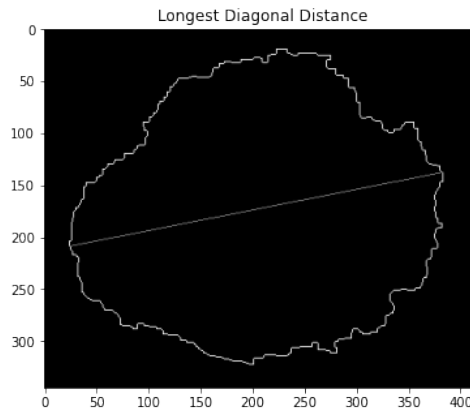


Figura 5: Mayor distancia diagonal del lunar.

2.4. Algoritmo para cada característica

Si bien en la sección 2.2 se definieron los 4 índices, no hay consenso preciso acerca de cómo proceder en su cálculo. Es decir que, y tomando como ejemplo al índice de asimetría (A), la definición indica que un lunar asimétrico obtendrá un valor de $A = 0$, uno que sea asimétrico a lo largo de solo uno de sus ejes le corresponderá valor $A = 1$, y si lo es en ambos entonces $A = 2$. Pero, ¿qué significa ‘ser asimétrico a lo largo de un eje’? No solo es cuestión de definir cómo medir dicha asimetría, sino que una vez hecho debe establecerse un umbral o punto de corte a partir del cual se considere ese resultado como de un lunar asimétrico o no. Esto debe repetirse para cada uno de los índices.

A continuación, nosotros proponemos una manera de calcular cada uno de los índices, sabiendo que existen otras y que cada una tendrá ventajas y desventajas.

2.4.1. Asimetría (A)

Se toma la imagen del lunar y se le calculan los ejes principales (se toma la mayor distancia diagonal y su recta mediatriz). De esta forma, cada uno de los ejes divide la lesión en dos mitades.

El siguiente paso es calcular el área (cantidad de píxeles blancos) que posee cada mitad, y compararla con la otra mitad. En caso de que la diferencia porcentual entre ambas sea mayor a un cierto umbral, entonces esas dos mitades se consideran suficientemente diferentes como para considerar que el lunar no es simétrico a lo largo de ese eje. Luego se repite para el otro eje.

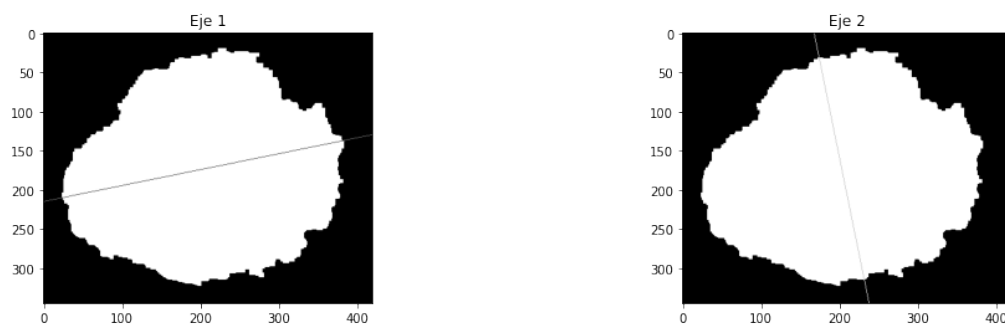


Figura 6: Obtención de los ejes para el cálculo de la asimetría

Luego de realizar diversas pruebas con las imágenes de la base de datos, se determinó que el umbral para esta métrica sea una diferencia porcentual del 17,5 % entre las áreas de cada mitad.

2.4.2. Bordes (B)

Se tomó la imagen del perímetro del lunar y se la dividió en 8 segmentos tomando como centro al centroide; tal como lo muestra la siguiente imagen:

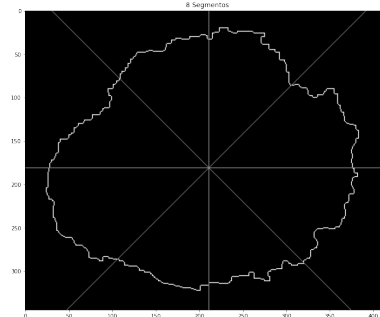


Figura 7: Obtención de los 8 segmentos.

Con esta división de la imagen, se armó un algoritmo que recorre cada porción comprendida entre los segmentos y se extraen los píxeles correspondientes al borde, generando así 8 imágenes que solo contienen los píxeles pertenecientes a uno de los segmentos. Para poder comprobar el correcto funcionamiento del algoritmo, se obtuvo la siguiente imagen:

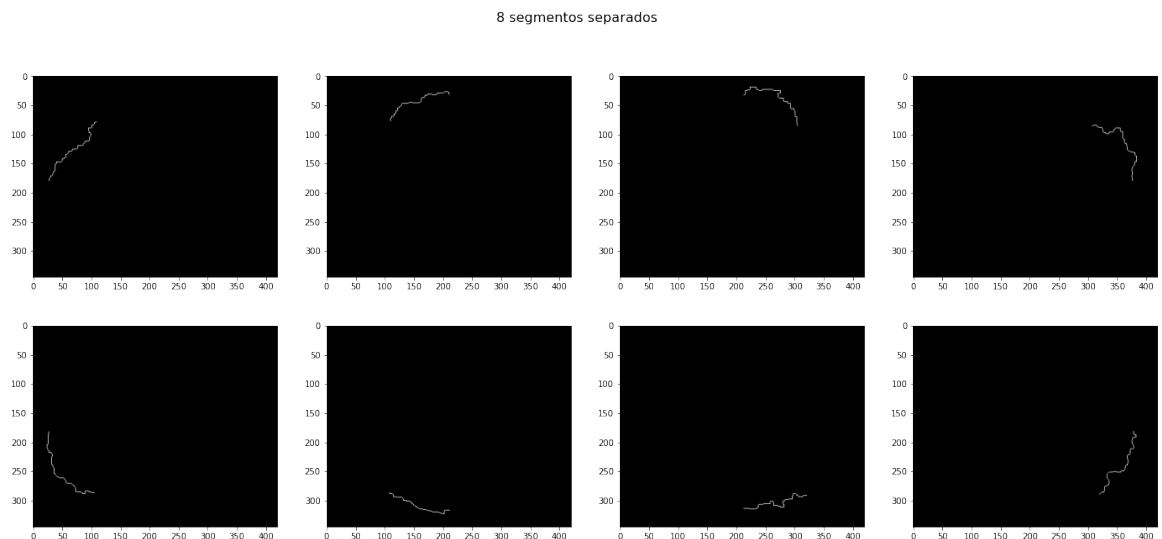


Figura 8: Obtención de los fragmentos de borde del lunar.

Finalmente, se calcularon parámetros estadísticos para observar si existen patrones de regularidad entre los fragmentos. Para ello, se calculó la distancia de cada píxel de fragmento de borde al centroide, y entre estas distancias se calculó la media, varianza, desvío y coeficiente de variación (CV). De todos estos, se determinó que el parámetro más relevante en este caso es el coeficiente de variación. Por lo tanto, el índice de Borde (B) se determinó con valores de 0 a 8 según la cantidad de fragmentos que tienen un coeficiente de variación superior al umbral establecido en 0,06.

2.4.3. Color (C)

Primero se definieron los valores RGB de los colores de referencia:

- Blanco: [255,255,255]
- Negro: [0,0,0]
- Rojo: [255,0,0]

- Marrón Claro: [200,150,100]
- Marrón Oscuro: [100,50,0]
- Gris Azulado: [100,125,150]

Se recorre la imagen píxel a píxel se calcula la distancia euclídea de ese color de píxel con la de todos los colores de referencia. Si esa distancia es menor al umbral específico para ese color, se considera que ese píxel es de ese color y se suma a un contador especificado para cada uno. Si la cantidad de píxeles del contador de cierto color, supera determinada proporción, entonces se considera que ese color se encuentra presente en el lunar. Este último paso de los contadores se realiza de modo de garantizar que exista una cantidad significativa de píxeles asociados a determinado color. Por ejemplo, sin estos contadores, con que haya 1 píxel que cumpla con el umbral de cierto color, entonces ya diríamos que ese color se encuentra en el lunar. Lo que terminaría ocurriendo es que todos los lunares analizados siempre tendrían un índice de $C = 6$, es decir, que presentan todos los colores. La proporción necesaria para los contadores de cada color se define en base a porcentajes del total de píxeles de la imagen los cuales fueron configurados en base a las imágenes de prueba que se utilizaron durante el desarrollo.

Para el cálculo de los umbrales, se siguió el procedimiento que realizó Ihab Zaquout [6]. Para ello, se utilizó la siguiente tabla de valores RGB:

| White | Black | Red | Light-brown | Dark-brown | Blue-gray |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| (1, 1, 1) | (0, 0, 0) | (1, 0, 0) | (0.7843, 0.5882, 0.3922) | (0.5882, 0.3922, 0.3922) | (0.5882, 0.4902, 0.5882) |
| (0.9608, 0.9608, 0.9608) | (0.0392, 0.0392, 0.0392) | (1, 0.1961, 0.1961) | (0.7843, 0.3922, 0) | (0.4902, 0.2941, 0.2941) | (0.4902, 0.4902, 0.5882) |
| (0.9216, 0.9216, 0.9216) | (0.0784, 0.0784, 0.0784) | (0.7843, 0, 0) | (0.7843, 0.3922, 0.1961) | (0.3922, 0.1961, 0.1961) | (0.3922, 0.3922, 0.4902) |
| (0.8824, 0.8824, 0.8824) | (0.1176, 0.1176, 0.1176) | (0.7843, 0.1961, 0.1961) | (0.5882, 0.3922, 0.1961) | (0.3922, 0.1961, 0) | (0.3922, 0.4902, 0.5882) |
| (0.8431, 0.8431, 0.8431) | (0.1569, 0.1569, 0.1569) | (0.5882, 0, 0) | (0.5882, 0.3922, 0) | (0.3922, 0, 0) | (0.1961, 0.3922, 0.5882) |
| (0.8039, 0.8039, 0.8039) | (0.1961, 0.1961, 0.1961) | (0.5882, 0.1961, 0.1961) | (0.5882, 0.1961, 0) | (0.1961, 0, 0) | (0, 0.3922, 0.5882) |

Figura 9: Tabla con valores RGB

Para el cálculo del umbral T_k correspondiente a cada color, se calcula como la distancia euclídea entre el máximo y el mínimo punto de referencia. Por ejemplo, el umbral para el color blanco, se calcularía como:

$$T_1 = \sqrt{3 \cdot (1 - 0.8039)^2} \quad (2)$$

Y el umbral del Marrón Claro, sería:

$$T_4 = \sqrt{(0.7843 - 0.5882)^2 + (0.5882 - 0.1961)^2 + (0.3922 - 0)^2} \quad (3)$$

Luego, al multiplicar estos umbrales por 255, ya obtenemos las distancias mínimas de los valores de los colores de los píxeles.

2.4.4. Diámetro (D)

Se calculó la distancia real en mm del diámetro del lunar (la mayor distancia diagonal). Esto se calcula mediante la siguiente fórmula:

$$M = \frac{25.4 \cdot diam}{20 \cdot dpi} \quad (4)$$

donde *diam* es el valor del diámetro en cantidad de píxeles, y dpi se corresponde con la cantidad de puntos por pulgada (que es un valor característico del dermatoscopio, y en este caso es de 96 puntos por pulgada). Se puede observar que el número 25.4 corresponde al factor de conversión de pulgadas a milímetros para obtener el valor M correspondiente al diámetro del lunar en milímetros.

En base a estos valores se asigna un valor de índice D a cada longitud del diámetro:

- $D = 0$, si $M < 1mm$
- $D = 0,5$, si $1mm < M < 2mm$
- $D = 1$, si $2mm < M < 3mm$
- $D = 1,5$, si $3mm < M < 4mm$
- $D = 2$, si $4mm < M < 5mm$
- $D = 2,5$, si $5mm < M < 6mm$
- $D = 3$, si $6mm < M < 7mm$
- $D = 3,5$, si $7mm < M < 8mm$
- $D = 4$, si $8mm < M < 9mm$
- $D = 4,5$, si $9mm < M < 10mm$
- $D = 5$, si $M > 10mm$

3. Resultados

Se probaron los algoritmos con imágenes de la base de datos PH2. El 50 % de ellas correspondientes a diagnósticos benignos y el resto correspondientes a diagnósticos como melanoma. Así, se calculó cada una de las métricas mencionadas anteriormente y se obtuvo el valor TDS usando la ecuación 1.

Se pudo observar que en un 83 % de los casos hubo un diagnóstico que coincidía con el anotado manualmente por médicos en la base de datos. Realizando un estudio de exactitud diagnóstica, encontramos que la sensibilidad (proporción de melanomas que tendrán test positivo) fue de 0.75; la especificidad (proporción de lunares benignos que tendrán test negativo) fue 1; el Valor Predictivo Positivo VPP (probabilidad de que el paciente tenga melanoma dado que el test es positivo) fue 1; y el Valor Predictivo Negativo VPN (probabilidad de que no sea melanoma dado que el diagnóstico es negativo) fue 0.75.

4. Conclusiones, limitaciones, y futuras mejoras

Si bien la interfaz gráfica goza de algunas limitaciones, la misma cumple su función como estimación temprana de lunares sospechosos.

La principal limitación importante está el cálculo de los umbrales, los cuales fueron elegidos manualmente y para usarse en las imágenes particulares con las que se trabajó. La manera correcta de elegir estos umbrales, es la que suelen usar en los distintos papers e investigaciones consultados: utilizando herramientas de Machine Learning, combinando los algoritmos utilizados con algoritmos de clasificación no supervisada, y así calcular automáticamente los umbrales óptimos. Esto requiere de conocimiento y habilidad en el mundo del Machine Learning, y además un gran número de imágenes para entrenar y luego evaluar al algoritmo. Una cantidad mayor a la que se puede analizar mediante un procedimiento ‘manual’ como el de este trabajo en un tiempo razonable.

Otro aspecto de mejora es utilizar algoritmos para eliminar los pelos presentes en la superficie de la piel. Existen casos en los que los pelos pueden estar en la superficie de los lunares, lo que puede llegar a perjudicar a la segmentación del lunar. Los pelos se pueden eliminar mediante algoritmos como la transformada Top-Hat que utiliza operaciones morfológicas para obtener imágenes que puedan segmentarse con mayor facilidad, tal como lo realizado por Jaworek et.al. [4]

Además, como se mencionó en la sección 2.4, los algoritmos implementados para el cálculo de los 4 índices fueron propuestas propias, pero estamos conscientes de que hay métodos más robustos y/o precisos para hacerlo.

Por último, también es muy importante hacer un comentario respecto a la segmentación y su posterior mejoramiento mediante operaciones morfológicas. Esto fue realizado de acuerdo a los conocimientos y habilidades obtenidas durante la cursada pero, si bien esto fue suficiente para segmentar correctamente los lunares benignos (como el de las imágenes previas), no se logró segmentar completamente bien a los melanomas. Como se ve en la figura 10, a pesar del mejoramiento mediante operaciones morfológicas, la segmentación de este melanoma presenta algunos agujeros y también ‘islas’ de píxeles alrededor. Esto interfiere en los cálculos básicos de área, perímetro, centroide, y mayor distancia diagonal (ver figura 11, la mayor distancia diagonal se da con una de las ‘islas’), lo cual a su vez genera problemas al calcular los índices (ver figura 12, donde los agujeros e ‘islas’ externas son tomados como píxeles de borde y por ende influyen en el cálculo de la media y varianza de sus distancias al centroide.)

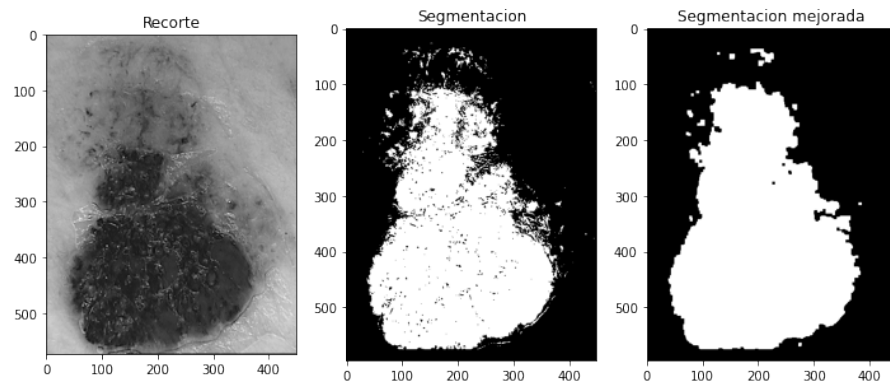


Figura 10: Notar que a pesar de la segmentación y su mejoramiento, no se logra un muy buen resultado.

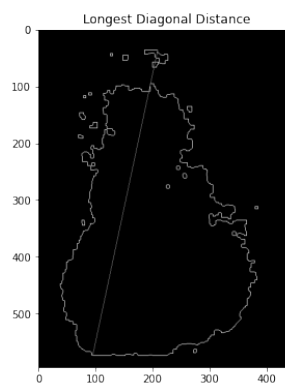


Figura 11: Mayor distancia diagonal del melanoma.

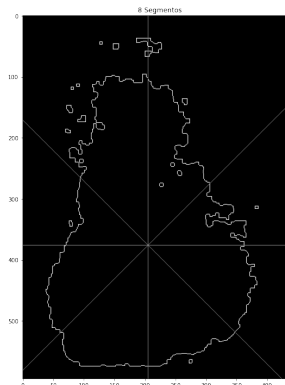


Figura 12: Obtención de los 8 segmentos para el melanoma.

5. Anexo: Capturas de la GUI

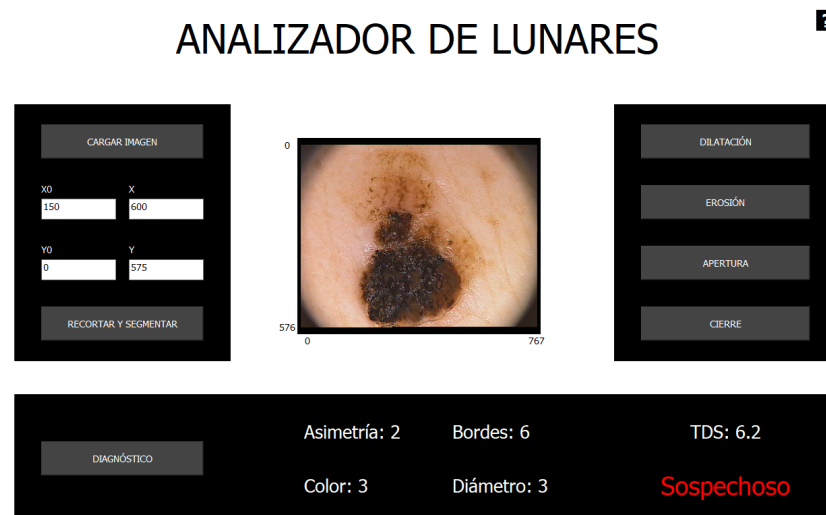


Figura 13: Interfaz Gráfica Desarrollada

6. Referencias

- [1] Medline Plus, "Lunares", disponible en <https://medlineplus.gov/spanish/moles.html>
- [2] American Cancer Society, "¿Qué es el cáncer de piel tipo melanoma?" disponible en: <https://www.cancer.org/es/cancer/cancer-de-piel-tipo-melanoma/acerca/que-es-melanoma.html>
- [3] Minkin, Elise; Keshwani, Sneha; Gadiraju, Abhishek; "Mole Investigator: Detecting Cancerous Skin Moles Through Computer Vision" University of Pennsylvania, 2014
- [4] Jaworek-Korjakowska, Joanna y Tadeusiewicz, Ryszard, "Hair removal from dermoscopic color images", Bio-Algorithms and Med-Systems, 2013
- [5] ADDI Project, "PH2 Database" disponible en: <https://www.fc.up.pt/addi/ph2%20database.html>
- [6] Ihab Zaqout, "Diagnosis of Skin Lesions Based on Dermoscopic Images Using Image Processing Techniques"