

Clasificación de Lesiones Cutáneas Utilizando Métodos de Procesamiento de Imágenes y Aprendizaje Profundo

Proyecto final de Grado, Bioingeniería

Autores: Choi, David y Migliano, Luciana

Supervisor: Dr. Milano, Federico

Co-Supervisor: Dr. Mosquera, Tomás

Departamento de Bioingeniería
Instituto Tecnológico de Buenos Aires

18 de octubre de 2019

Resumen

El melanoma es uno de los tipos de tumores malignos más letales, representando el 79 % de las muertes de los pacientes con cáncer de piel. La detección temprana de esta patología es crucial ya que la tasa de supervivencia de los pacientes depende del estadio en el que se encuentra el melanoma al ser diagnosticado, variando entre un 15 % y 65 % en el estadio terminal y temprano, respectivamente. Hoy en día el diagnóstico clínico del melanoma depende completamente de la subjetividad del médico que lo evalúa visualmente.

El reconocimiento automático de los melanomas es una tarea desafiante ya que existe una gran variedad de artefactos en las imágenes y mucha variabilidad en la morfología de las lesiones. El procesamiento de las imágenes y las redes neuronales profundas se convierten en protagonistas por el gran potencial que tienen para reconocer y clasificar lesiones sospechas con una gran precisión, facilitando así la tarea de los médicos y favoreciendo el diagnóstico temprano de melanomas.

Un dermatólogo experto analizó 124 imágenes dermatoscópicas e identificó tres tipos de patrones (velo azul blanquecino, red pigmentada atípica y asimetría) en cada una de ellas y las clasificó en dos tipos (sospechosas o no sospechosas de ser una lesión melanocítica). Se desarrolló un software de procesamiento de imágenes para identificar estos tres criterios dermatoscópicos automáticamente, utilizando como *Gold Standard* la clasificación del experto. Por otro lado, se entrenó una red neuronal convolucional de arquitectura ResNet50 con 4000 imágenes dermatoscópicas y sus resultados histopatológicos (2000 benignas y 2000 melanomas). Luego se evaluaron las predicciones de la red neuronal en 604 imágenes pertenecientes a la base de datos de evaluación con los resultados histopatológicos de las lesiones.

Se obtuvo una precisión de 72.00 % y sensibilidad de 81.2 % en el método basado en procesamiento de imágenes; y una precisión de 78.15 % y sensibilidad de 89.90 % en el método basado en inteligencia artificial. Los métodos propuestos muestran resultados buenos comparado con los métodos existentes para la clasificación de lesiones cutáneas.

Índice

1. Introducción	2
1.1. Consideraciones dermatológicas	3
1.1.1. Piel [1]	3
1.1.2. Melanoma	5
1.2. Métodos de <i>screening</i>	5
1.2.1. Método ABCDE	5
1.2.2. Método de Menzies	6
1.2.3. Método de los 7 puntos de Argenziano [2]	6
1.2.4. Método de la Regla de los 3 Puntos	7
2. Antecedentes	9
2.1. Detección de los patrones	9
2.1.1. Velo azul blanquecino	9
2.1.2. Red de pigmento atípica	10
2.1.3. Asimetría	11
2.2. Redes neuronales	12
3. Objetivos	13
3.1. Objetivo general	13
3.2. Objetivos específicos	13
4. Materiales y métodos	14
4.1. Dermatoscopía	14
4.2. Base de datos	17
4.3. Método de procesamiento de imágenes	19
4.3.1. Preprocesamiento de las imágenes	20
4.3.2. Segmentación	21
4.3.3. Detección de patrones	22
4.3.4. Criterio de clasificación	25
4.4. Método de inteligencia artificial	25
4.4.1. Método de redes neuronales	26
4.4.2. Preprocesamiento	26
4.4.3. <i>Transfer Learning</i>	31
4.5. Métricas de evaluación	37
5. Resultados	39
5.1. Método de procesamiento de imágenes	39
5.1.1. Resultados globales	39

5.1.2. Resultados de segmentación	40
5.1.3. Resultados de cada uno de los patrones detectados	40
5.2. Método de inteligencia artificial	42
5.3. Comparación de métodos	45
6. Discusión	47
6.1. Método de procesamiento de imágenes	47
6.1.1. Problemas de segmentación	48
6.1.2. Selección del método dermatoscópico a implementar	48
6.1.3. Espacio de color	49
6.1.4. Resolución de las imágenes	49
6.2. Método de inteligencia artificial	50
6.2.1. Segmentación en el preprocesamiento de las imágenes	51
6.2.2. Composición del conjunto de entrenamiento	52
6.2.3. Cantidad de imágenes del conjunto de entrenamiento	52
6.2.4. Considerar los metadatos para el entrenamiento de la red neuronal .	52
6.3. Comparación de métodos	52
7. Conclusiones	54
A. Apéndice	55
Referencias	59

1. Introducción

Según estadísticas globales, uno de cada tres diagnósticos de cáncer corresponden a cáncer de piel [3], y se estima que uno de cada cinco estadounidenses desarrollarán este tipo de cáncer en algún momento a lo largo de su vida [4]. A su vez, la continua reducción de los niveles de ozono en la atmósfera dificulta la filtración de los rayos ultravioletas (UV), aumentando así la exposición de las personas y el riesgo de padecer cáncer de piel. En Argentina, se atribuye el 52.80 % de los casos de melanoma a la exposición a rayos ultravioletas [5] y mueren más de 10 personas por semana debido a este tipo de cáncer (La probabilidad de sobrevivir a un melanoma a cinco años si es detectado en los primeros estadios del tumor es de alrededor del 97 %, mientras que si es detectado en estadios más avanzados, se puede reducir la probabilidad hasta un 14 %). A esto se le suma que de dos a tres millones de personas en el mundo son diagnosticadas anualmente de cáncer de piel y que el estadio del cáncer al ser detectado es inversamente proporcional a la probabilidad de supervivencia, por lo que es de vital importancia desarrollar métodos de detección temprana de esta enfermedad.

La incidencia de los melanomas sigue aumentando a pesar de los esfuerzos de gobiernos y asociaciones dermatológicas de distintos países en cuanto a la concientización de los daños que produce la exposición solar y de las recomendaciones de realizar chequeos dermatológicos periódicos. [6]

La metodología utilizada actualmente para la detección de cáncer de piel consiste en un examen de *screening* que lo lleva a cabo un dermatólogo. Los dermatólogos utilizan distintas técnicas para evaluar las características de las lesiones cutáneas, siendo las más utilizadas el Método de ABCD [7], el Método de Menzies [8], el Método de Argenziano [9] y la Regla de los 3 puntos [2]. Estos métodos permiten analizar las lesiones de una persona en busca de determinar si alguna de éstas requiere de un análisis más profundo, y al ser sencillos permiten que puedan ser utilizadas por profesionales inexpertos. Estos criterios se basan en un sistema de puntos para decidir si la lesión bajo observación se trata de una lesión sospechosa o no sospechosa. Seguido de esto, se realiza una biopsia de los lunares o manchas sospechosas para confirmar el diagnóstico presuntivo y definir la patología de la lesión.

Para conseguir un aumento de la precisión y así reducir la cantidad de extracción de tejido para análisis histopatológico, se desarrollaron métodos más sofisticados y complejos de diagnóstico como la dermatoscopia [10]. Ésta es una técnica *in vivo*, no invasiva, que permite identificar estructuras de la piel que no se distinguen a simple vista. Es una técnica sencilla, rápida y económica que consiste en la observación de la lesión mediante una lente de 10 a 30 aumentos con una fuente de iluminación potente. Para evitar la dispersión de la luz en contacto con la capa córnea, los dispositivos suelen utilizar luz polarizada o luz convencional pero aplicando aceite de inmersión sobre la piel y contactando directamente

la lente a la superficie cutánea [11]. Esta técnica permitió aumentar la precisión diagnóstica de melanomas en profesionales entrenados hasta entre un 75 % y 80 % [12].

El desarrollo de un software que detecte automáticamente el riesgo de una lesión en la piel a partir de los resultados histopatológicos de una gran cantidad de lesiones sería de mucha utilidad ya que reduciría significativamente el tiempo de análisis de imágenes de los médicos y hasta podría aumentar la precisión diagnóstica con la que se cuenta mediante el empleo de los métodos tradicionales.

En los últimos años la inteligencia artificial aplicada al campo de la salud fue brindando soluciones que abarcan desde la detección de patrones para predecir el deterioro de la salud de un paciente hasta redes neuronales que identifican patologías en distintos tipos de imágenes biomédicas. Según un reporte de Frost & Sullivan [13], el apoyo clínico de la inteligencia artificial fortalecerá los procesos de diagnóstico donde intervienen las imágenes médicas. Incluso se estima que esta tecnología tiene el potencial de mejorar los resultados entre un 30 % y un 40 %. Friedman et al. [14] llevaron a cabo un estudio sobre el desempeño de dermatoscopistas en el diagnóstico de lesiones pigmentadas cutáneas pequeñas (con diámetro inferior a 6 mm.) y lo compararon con un sistema automático de visión artificial, y obtuvieron como resultado que el sistema artificial fue más efectivo.

El objetivo de este trabajo no es reemplazar el diagnóstico por los profesionales médicos, sino ayudarlos en la toma de decisiones al contar con más herramientas y con más información. En este trabajo se propone el desarrollo de dos métodos de clasificación automática basado en dos técnicas: procesamiento de imágenes médicas tradicionales y aprendizaje profundo a través de redes neuronales. Así, para mejorar la precisión diagnóstica de los melanomas se combinan técnicas de dermatología con técnicas de ingeniería, como el procesamiento de las imágenes y la inteligencia artificial.

Este trabajo fue implementado en el lenguaje de programación *Python* [15]. Se utilizó una base de datos de un total de 4124 imágenes dermatoscópicas (124 imágenes para la clasificación con métodos de procesamiento de imágenes y 4000 imágenes para la red neuronal), etiquetadas con el resultado histopatológico de cada lesión. Se desarrollaron dos métodos diferentes. Por un lado, el procesamiento de imágenes tradicionales se realizó en cuatro etapas, que consistieron en preprocesamiento, segmentación, detección de patrones y clasificación de imágenes, tal como se observa en la figura 1.

Por otro lado, se utilizó *transfer learning* [16] para entrenar una red neuronal para lograr predecir el diagnóstico de las lesiones, clasificadas en benignas o melanomas.

1.1. Consideraciones dermatológicas

1.1.1. Piel [1]

La piel es el órgano que recubre toda la superficie externa del cuerpo y uno de los más importantes por su tamaño y funciones. Por su estructura, se compone de tres capas: la

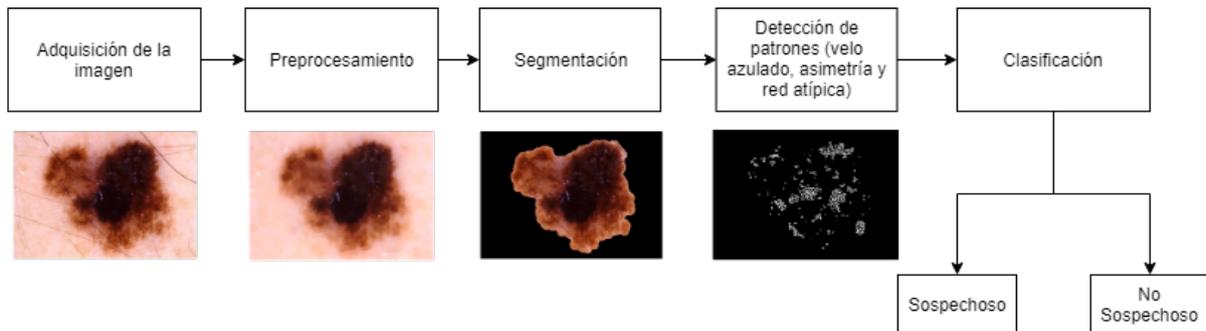


Figura 1: Diagrama de bloques del programa de clasificación con métodos de procesamiento de imágenes.

dermis, la epidermis y el tejido celular subcutáneo (o hipodermis). La figura 2 muestra las capas de la piel junto con las células más importantes que componen la epidermis.

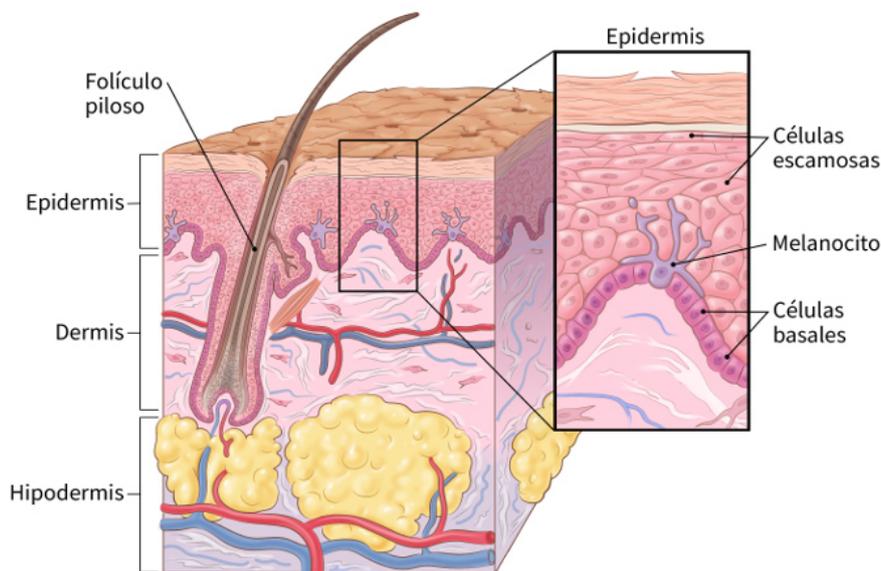


Figura 2: Representación esquemática de las capas de la piel normal.

La epidermis es un epitelio superficial y es la capa más externa de la piel, por lo que su función principal es la protección de acciones lesivas del medio y la pérdida de líquidos. Las células principales de esta capa son los queratinocitos y constituyen el 80 % de las células epidérmicas. También se encuentran los melanocitos, que son los que le dan la pigmentación a la piel, y las células de Langerhans y linfocitos, que son las encargadas de las defensas inmunitarias cutáneas y, por lo tanto, primera barrera de defensa del organismo.

La dermis es la capa gruesa de tejido conectivo a la que se fija la epidermis. En esta capa están incluidos los folículos pilosos y las glándulas sudoríparas. Las células que se encuentran principalmente en la dermis son fibroblastos y macrófagos, pero también hay gran cantidad de mastocitos. Esta capa está compuesta por una gran cantidad de colágeno.

Por último, el tejido celular subcutáneo - o hipodermis -, que suele contener abundancia

de lípidos, relaciona la piel con las estructuras más profundas. Esta capa cumple la función de depósito de energía y de aislante para mantener el calor corporal.

1.1.2. Melanoma

El melanoma es una neoplasia maligna generada en los melanocitos de la capa basal de la epidermis desde donde se extiende a las capas más superficiales, a la dermis y al tejido celular subcutáneo, con alta capacidad para producir metástasis.

Es la forma más peligrosa del cáncer de piel, ya que aunque representa solo un 4% de todos los cánceres dermatológicos, es responsable del casi 90% de las muertes [17]. La tasa de supervivencia a 5 años de los pacientes con melanoma metastásico es solamente de un 14% [18]. Esta enfermedad fue incrementando su incidencia en las últimas décadas más que cualquier otro tipo de cáncer. En los últimos 30 años, la incidencia del melanoma en la población de los Estados Unidos aumentó tres veces, estimándose que uno de cada 36 hombres y una de cada 55 mujeres estadounidenses desarrollarán algún tipo de melanoma a lo largo de su vida [19].

Actualmente, ningún tipo de cáncer de piel es completamente prevenible. Sin embargo, se sabe que su origen es multifactorial y que la principal causa desencadenante es la exposición a la radiación ultravioleta del sol intensa de tipo intermitente. Hay evidencias epidemiológicas que muestran que en los países donde se implementaron estrategias de protección a la exposición solar décadas atrás, disminuyeron los melanomas y muertes asociadas al mismo [20] [21].

La prevención primaria consiste en disminuir la exposición a luz ultravioleta para que el riesgo de desarrollar melanoma sea menor. La prevención secundaria consiste en el diagnóstico temprano y el tratamiento oportuno de los melanomas que llevan a una reducción de su morbilidad y mortalidad [19].

A fin de promover el diagnóstico temprano y el tratamiento de melanoma, se han implementado exámenes de *screening* de la piel, capacitaciones sobre la apariencia clínica de la enfermedad y nuevos dispositivos para profesionales.

1.2. Métodos de *screening*

1.2.1. Método ABCDE

La técnica más utilizada para un análisis clínico primario es la regla ABCD, cuyas iniciales corresponden a asimetría, bordes irregulares, colores múltiples y diámetro mayor a 6 mm. Este modelo fue desarrollado en 1985 por Friedman et al. [7] para poder identificar lesiones basándose en cuatro criterios fáciles de recordar, aprender y aplicar para observadores poco experimentados. Para aumentar la precisión del diagnóstico, luego se incorporó un quinto criterio que es la evolución (representada por la E en el nombre

del método). Se considera una lesión sospechosa si uno o más de estos patrones están presentes en su morfología.

Este método diagnóstico tiene una tasa de exactitud baja (entre 50% y 75%). Esto se debe en parte a que los melanomas en las fases iniciales son muy difíciles de identificar porque comparte muchas características clínicas con un nevus atípico [22]. Además, los melanomas en las fases iniciales tienen un diámetro menor a lo que indica la regla (6 mm), por lo que el criterio de ABCDE no es de gran utilidad para la detección temprana de este tipo de cáncer. [23]

1.2.2. Método de Menzies

El método creado por Menzies et al. [8] en 1966 también para simplificar el diagnóstico de melanomas por profesionales inexpertos, considera 11 criterios dermatoscópicos, divididos en criterios negativos y positivos. Para que una lesión se considere melanoma, no debe presentar ninguno de los dos rasgos negativos y debe presentar uno o más rasgos positivos. En el cuadro 1 se muestran los rasgos analizados en este método.

Rasgos negativos	Rasgos positivos
Simetría del patrón de pigmentación Un solo color	Velo blanco azulado Prolongaciones radiales Pseudópodos Puntos/glóbulos negros periféricos Múltiples puntos marrones Múltiples puntos azules grisáceos Áreas blanquecina tipo cicatriz Colores múltiples Red de pigmento ensanchada

Cuadro 1: Tabla de los rasgos negativos y positivos del método de Menzies.

Este método presenta una sensibilidad superior a otros métodos de diagnóstico de melanomas.

1.2.3. Método de los 7 puntos de Argenziano [2]

Este método de diagnóstico semicuantitativo creado por Argenziano et al. [9] considera 7 criterios dermatoscópicos. El sistema de puntaje cuantitativo permite una simplificación del análisis de patrones clásicos, ya que se abordan menos rasgos. El primer paso consiste en clasificar la lesión en melanocítica o no melanocítica. Luego se aplica este procedimiento si la lesión se clasifica como melanocítica. Los patrones que se consideran son los siguientes:

- Red de Pigmento Atípica
- Velo azul blanquecino

- Patrón vascular atípico
- Líneas o proyecciones irregulares
- Puntos/Glóbulos irregulares
- Manchas irregulares
- Regresión

La clasificación se define por un sistema de puntuación asignado a cada rasgo. La lesión es considerada maligna si la puntuación total supera los 3 puntos.

En el cuadro 2 se presenta el sistema de puntaje utilizado para la clasificación:

Criterio dermatoscópico	Puntuación máxima
CRITERIOS MAYORES	
Retículo pigmentado atípico	2
Velo azul-blanquecino	2
Patrón vascular atípico	2
CRITERIOS MENORES	
Proyecciones irregulares	1
Puntos/glóbulos irregulares	1
Manchas de pigmento irregulares	1
Estructuras asociadas a regresión	1
PUNTUACIÓN TOTAL	
<3	≥ 3
BENIGNA	MALIGNA

Cuadro 2: Lista de los 7 puntos para el diagnóstico entre lesiones melanocíticas benignas y malignas.

En el estudio [9] se obtuvo que un puntaje de 3 o más clasifica a una lesión como melanoma, con una sensibilidad de 95 % y una especificidad de 75 %. En total, el 82 % de las lesiones melanocíticas fueron diagnosticadas correctamente con este método. Esta regla permite obtener la mayor sensibilidad para los profesionales inexpertos.

1.2.4. Método de la Regla de los 3 Puntos

En la Reunión para el Consenso sobre Dermatoscopia en el 2001, se definió que tres criterios eran particularmente importantes para diferenciar lesiones pigmentadas cutáneas malignas de las benignas [2]. Estos tres criterios son la presencia de asimetría, red pigmentada atípica y velo azul blanquecino. Estudios preliminares demostraron que la presencia de dos de estos tres criterios, están asociados con una alta probabilidad de que la lesión sea maligna, por lo que la presencia de más de un criterio sugiere una lesión sospechosa.

En un principio se diseñó este método para un análisis de patrones dermatoscópicos simplificados, a ser utilizado por médicos generales y para evitar ignorar un melanoma, por lo que es más sensible que específico.

Un estudio abierto de internet [24] estimó la validez del método con un grupo de médicos inexpertos y obtuvo los siguientes resultados:

- La sensibilidad para lesiones malignas fue del 91 % y no estuvo influenciado por la experiencia de los observadores.
- La especificidad fue del 71.9 % y fue influenciada significativamente por la profesión, siendo los dermatólogos los que obtuvieron el mejor desempeño.

Este estudio concluyó que es un método simple, factible, preciso y reproducible para la detección temprana de cáncer de piel. Al utilizar este método, se alcanza una sensibilidad y especificidad comparables a las de otros métodos que requieren una formación muy superior.

Para el presente trabajo se utilizaron los criterios de este método para clasificar las lesiones cutáneas a través del procesamiento tradicional de las imágenes.

2. Antecedentes

Debido al interés que despertó en los últimos años el cáncer de piel - particularmente el melanoma - y a los avances en modelos automatizados para la clasificación de imágenes y de aprendizaje profundo, muchos grupos de investigación comenzaron la búsqueda de soluciones de *software* para identificar lesiones cancerígenas en imágenes dermatoscópicas. Se han desarrollado diversos enfoques para resolver los mismos problemas, por lo que a continuación se presentan algunos de los métodos publicados más importantes para identificar cada uno de los tres patrones estudiados en este trabajo así como también de los métodos de inteligencia artificial publicados para la clasificación de lesiones cutáneas. Los mismos se citarán a continuación.

2.1. Detección de los patrones

En este trabajo se analizará la detección del velo azul blanquecino, la presencia de red pigmentada atípica y se examinará la asimetría de las lesiones. A continuación se presentan los trabajos publicados más relevantes para la detección de cada patrón.

2.1.1. Velo azul blanquecino

El velo azul blanquecino es uno de los patrones más relevantes en la identificación de melanomas ya que es uno de los más específicos de este tipo de lesión, aunque también puede encontrarse en nevus fusocelulares o en lesiones combinadas [12]. Es una zona azul y sin estructuras focales con una capa tipo "vidrio esmerilado" blanca superpuesta [25]. La pigmentación azul no puede ocupar toda la lesión y generalmente corresponde a áreas elevadas o palpables de la misma. En la figura 3 se muestra un ejemplo de una lesión con presencia de este patrón.

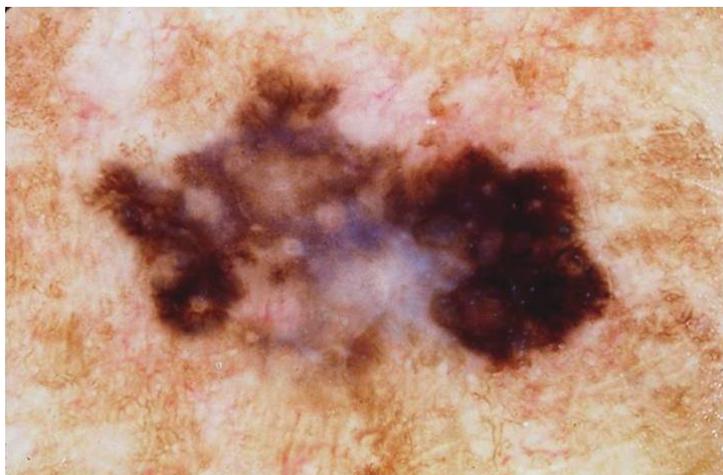


Figura 3: Lesión con presencia de velo azul blanquecino. Este patrón se puede observar en el centro de la lesión como una estructura de color azulada y blanquecino.

Se han investigado y desarrollado algunos métodos para identificar estas zonas en las lesiones cutáneas, siendo las más relevantes para este trabajo las que se mencionan a continuación.

En el equipo de Celebi et al. [26] utilizaron un método de aprendizaje automático para la detección del velo azul blanquecino en donde clasifican los píxeles de la imagen usando un árbol de decisión como clasificador. Obtuvieron una sensibilidad del 69.35 % y una especificidad del 89.97 % en la detección de áreas azul blanquecinas en un conjunto de 545 imágenes dermatoscópicas.

Madooei et al. [27] presentaron un método para detectar y segmentar automáticamente las áreas de velo azul blanquecino en imágenes dermatoscópicas a través del análisis de color y técnicas de visión artificial.

Por último en 2011, García Arroyo et al. [28] propusieron un método asistido por computadora para la detección de este patrón, obteniendo un resultado de imágenes detectadas correctamente del 89.06 %, en un total de 887 imágenes.

2.1.2. Red de pigmento atípica

El patrón de red de pigmento atípica es uno de los indicios presentes en el diagnóstico de melanoma. Aunque el patrón por sí solo no es indicador de malignidad, su hallazgo en conjunto con otros patrones indican que la lesión es sospechosa. Este patrón consiste en líneas rectas (estrechas o gruesas), distribuidas de manera tal que se cruzan entre sí en ángulos cercanos a los 90 grados, en intervalos regulares. Estas líneas corresponden a las crestas epiteliales ensanchadas llenas de melanocitos atípicos. La figura 4 muestra un ejemplo de una lesión que presenta red de pigmento atípica.

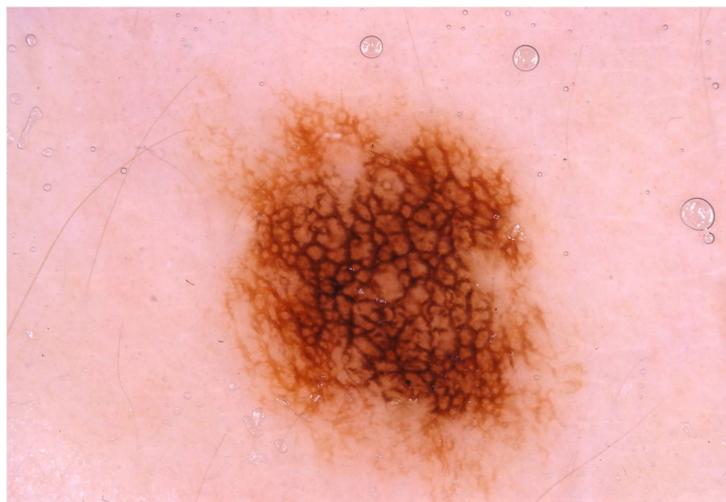


Figura 4: Lesión con presencia de red de pigmento atípica. La red se caracteriza por tener los bordes pigmentados y la superficie no encerrada. Las redes de pigmento se pueden encontrar en cualquier parte de la lesión, tanto como en el centro como en la periferia.

Una revisión del estado del arte para el reconocimiento de este patrón se llevó a cabo

por García Arroyo et al. [28]

En 2014 [29], desarrollaron un algoritmo de detección de redes pigmentadas en imágenes dermatoscópicas basados en aprendizaje automático supervisado y en análisis estructural. Probaron este método en una base de datos de 220 imágenes y obtuvieron un 86.00 % de sensibilidad y un 81.67 % de especificidad.

Luego, en 2018, el mismo grupo de investigación formado por García Arroyo et al. [30], desarrolló un método en base a la clasificación difusa de píxeles según tres categorías presentes en la definición de este patrón (red, agujero u otros). Los resultados mostraron una sensibilidad del 90.71 % y una especificidad del 83.44 % en un conjunto de 875 imágenes.

Eltayef et al. [31] presentaron un sistema automático para la detección de red pigmentada en imágenes dermatoscópicas. Al probarlo en una base de datos de 200 imágenes, obtuvieron una sensibilidad del 92.30 % y una especificidad del 95.00 %.

2.1.3. Asimetría

La asimetría se puede definir como una distribución discontinua de estructuras y colores, sin un patrón específico. Un ejemplo de este patrón se presenta en la figura 5.



Figura 5: Lesión con presencia de asimetría. No se puede definir ningún eje en la lesión que separe simétricamente ni las estructuras ni los colores.

En el año 1992, Stoecker et al. [32] propusieron un método de detección automática de asimetría en tumores de piel. Desarrollaron un algoritmo de segmentación, determinación de los ejes principales y de segmentación, obteniendo una precisión del 93.50 %, en una base de 46 imágenes.

Seidenari et al. [33], en el año 2006, desarrollaron un método basado en la división de la imagen en grillas, determinación de los ejes principales de la lesión y el cálculo de cada grilla a los ejes. De una base de 459 imágenes, obtuvieron una precisión de 92.40 %, una sensibilidad de 94.50 % y una especificidad del 90.20 %.

2.2. Redes neuronales

Los esfuerzos por encontrar soluciones automáticas para la clasificación de lesiones de piel utilizando inteligencia artificial fueron creciendo durante los últimos años. A medida que fueron avanzando los conocimientos en este campo y con el surgimiento de las redes convolucionales profundas, se empezaron a lograr mejores resultados en la clasificación de imágenes. A continuación se presentan algunos de los trabajos con resultados más destacados en esta área de investigación.

En 2004, Blum et al. [34] desarrollaron un algoritmo computacional para el diagnóstico de lesiones melanocíticas y compararon su precisión diagnóstica con los resultados de otras reglas de clasificación dermatoscópicas. Lograron una precisión diagnóstica del 82.00 % para imágenes completas y 84.00 % para lesiones con imágenes parciales. Encontraron que esta precisión se encontraba en el mismo rango de precisión diagnóstica que la aplicación de reglas de clasificación dermatoscópicas por expertos.

Un año después, Stanganelli et al. [35] introdujeron un diagnóstico asistido por computadora para mejorar la evaluación clínica de lesiones pigmentadas cutáneas, alcanzando una sensibilidad del 80.00 % y una especificidad del 79.77 % en un total de 477 imágenes.

Tomatis et al. [36] evaluaron el rendimiento de un nuevo sistema espectroscópico para el diagnóstico de los melanomas. De las imágenes extrajeron diferentes descriptores de lesiones relacionados con la distribución del color y la morfología de las lesiones. Luego aplicaron técnicas de reducción de datos, antes de crear un clasificador de una red neuronal para automatizar los diagnósticos. La red neuronal pudo discriminar entre melanomas y no melanomas con una sensibilidad del 80.40 % y una especificidad del 75.60 % en un conjunto de 1391 imágenes. Luego agregaron un set de 1022 lesiones y concluyeron que el rendimiento de su sistema automatizado es muy dependiente de la población.

En 2008, Iyatomi et al. [37] presentaron un sistema de screening de melanoma basado en un servidor web. Éste separa el área del tumor de la piel que lo rodea, calcula 428 características de la lesión para caracterizar el tumor, clasifica el mismo como melanoma o nevus utilizando un clasificador de red neuronal y presenta el diagnóstico. Su sistema alcanza una sensibilidad del 85.90 % y una especificidad del 86.00 % en un conjunto de 1258 imágenes dermatoscópicas utilizando cross-validation.

Por último en 2016, Yu et al. [38] propusieron un método para el reconocimiento de melanoma, aprovechando las redes neuronales profundas convolucionales. Utilizaron una red de más de 50 capas y dos pasos para mejorar el rendimiento. Lograron una especificidad de 93.10 % pero con una sensibilidad de 54.70 %, dando como resultado una precisión de 85.50 %.

3. Objetivos

3.1. Objetivo general

Desarrollar métodos de detección precoz de melanoma mediante el uso de un sistema automatizado.

3.2. Objetivos específicos

- Evaluar el desempeño de un método de procesamiento de imágenes dermatoscópicas de lesiones cutáneas, basado en el análisis de distintos patrones según la regla de los 3 puntos- velo azul blanquecino, red pigmentada atípica y asimetría- como indicadores de melanoma, contrastando el diagnóstico con el juicio de expertos como estándar de oro (*Gold Standard*).
- Evaluar el desempeño de un método de inteligencia artificial para clasificar las imágenes dermatoscópicas de lesiones cutáneas en melanoma o no melanoma, contrastando con el resultado histopatológico como *Ground Truth*.
- Comparar la efectividad de ambos métodos para la detección de melanomas.

4. Materiales y métodos¹

4.1. Dermatoscopía

La dermatoscopía, también llamada microscopía de epiluminiscencia, es una técnica no invasiva para examinar lesiones pigmentadas cutáneas utilizando un dermatoscopio, aunque también se utiliza para observar lesiones con poco o sin pigmento. Es particularmente útil para la identificación de los melanomas cutáneos malignos en una fase temprana. Con esta técnica se pueden visualizar mejor las estructuras pigmentarias o vasculares de la epidermis o dermis superficial, permitiendo un diagnóstico precoz de las lesiones potencialmente malignas.

El principio básico de la dermatoscopía es la transiluminación de una lesión y el posterior estudio con una lente de alto aumento para visualizar las características de la misma. Cuando la luz incide en la piel se pueden producir cuatro fenómenos físicos: absorción, difracción, refracción y reflexión. La absorción de la luz ocurre cuando la onda no atraviesa ni se refleja en un medio. La difracción es el fenómeno que hace que las ondas se desvíen al llegar a un medio o al atravesar una rendija. La refracción genera un cambio en la dirección y velocidad de una onda al atravesar una interfase entre dos medios distintos donde se propaga. Por último, la reflexión se basa en una modificación en la dirección de la onda cuando se encuentra con otro medio distinto al medio en el que se está propagando, generando una desviación de la misma hacia su medio original, evitando que la onda atraviese el otro medio. La figura 6 muestra estos fenómenos.

Estos principios son importantes ya que la mayor parte de la luz que incide en la piel seca se refleja, pero si se la moja con aceite, la mayor parte de la luz atraviesa la piel hasta la dermis profunda, permitiendo observar las estructuras de éstas. Este fenómeno permitió que los primeros dermatoscopios puedan hacer que la luz alcance las capas profundas de la piel y que las estructuras debajo de las lesiones de la piel puedan ser vistas. Estos dispositivos cuentan con tres componentes esenciales [39]:

1. *Lente acromática*: generalmente se utilizan objetivos de aumento 10x.
2. *Fuente de luz integrada*: se suelen utilizar fuentes de luz orientadas con un ángulo de 20° en los dermatoscopios manuales. Se pueden utilizar distintos tipos de fuentes de luz. Las lámparas halógenas se solían utilizar pero el color amarillo de la lámpara altera los colores que se visualizan de la piel. Una alternativa son los diodos emisores de luz (LED por sus siglas en inglés) que proveen luz blanca de alta intensidad y consumen un 70 % menos de potencia que las lámparas de halógeno. También existen dispositivos que emiten luces de diferentes colores ya que la penetración de la luz en

¹El código de los métodos discutidos en esta sección se encuentran publicados en <https://github.com/lumigliano/deteccion-melanoma>.

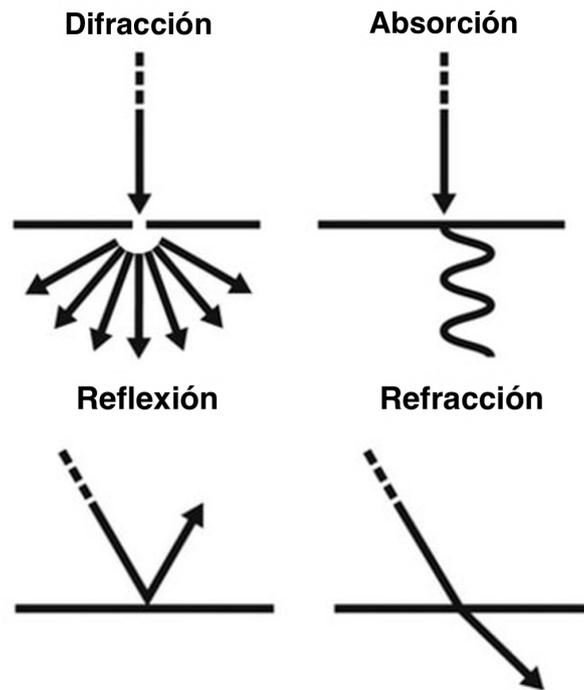


Figura 6: Efectos físicos de una onda lumínica.

la piel depende de la longitud de onda de la misma, por lo que mejora la visualización de la piel.

3. Fuente de alimentación

La figura 7 muestra un dermatoscopio manual comercial de la marca Dermlite®.



Figura 7: Dermoscopio manual de la marca Dermlite®.

Algunos dispositivos además cuentan con un sistema de fotografía para capturar y almacenar las fotos de las lesiones. La figura 8 muestra un dispositivo de este estilo.



Figura 8: Dermatoscopio VEOS® SLR con una cámara Canon integrada.

La dermatoscopia fue avanzando con el correr de los años y aunque al principio se utilizaban dermatoscopios de contacto (donde el vidrio del instrumento está en contacto con los fluidos que se le aplican a la superficie de la lesión), desde hace aproximadamente una década se comenzaron a utilizar los dermatoscopios de luz polarizada. Éstos utilizan distintos tipos de filtros luminosos para obtener luz cruzada polarizada, de forma de disminuir la reflexión y refracción de la luz incidente sobre la superficie de la piel. Estos tipos de dispositivos suelen incorporar unos separadores que sirven para distanciar la lente de la superficie de la piel a una distancia óptima para obtener una mejor visualización de la lesión. Según se trate de un dermatoscopio de contacto o de luz polarizada, se utiliza de una manera u otra. Para los dispositivos de contacto, primero se aplica el fluido de inmersión sobre la lesión a estudiar. Luego se sumerge el dermatoscopio en el fluido y se presiona sobre la piel hasta que desaparezcan las burbujas en el líquido. Por último se ajusta la distancia del dermatoscopio para obtener una imagen nítida de la lesión. Por otro lado, para los dermatoscopios de luz polarizada sólo se necesita acercarse al dispositivo hasta aproximadamente 1 cm de la piel (para aquellos sin separadores) o si cuenta con un separador, se apoya directamente sobre la lesión. Luego se enciende la fuente de iluminación y si no se obtiene una imagen nítida de la lesión, se ajusta la distancia acercándolo o alejándolo de la superficie cutánea hasta obtener una buena imagen. La dermatoscopia se utiliza cada vez más ya que tiene una serie de ventajas competitivas con respecto al método tradicional de ABCDE. Primero, aumenta significativamente (alrededor del 10 % a 27 %) la precisión del diagnóstico in vivo de los melanomas. También puede diferenciar la mayoría de las lesiones cutáneas de los melanomas, reduciendo la cantidad de biopsias innecesarias en los pacientes.

A pesar de mejorar la precisión del diagnóstico, la dermatoscopia no puede hacer un diagnóstico completamente certero de las lesiones, por lo que no puede sustituir totalmente a los estudios histopatológicos que se les hacen luego de ser extirpadas o biopsiadas.

4.2. Base de datos

Las imágenes utilizadas corresponden a la base de datos de *International Skin Imaging Collaboration* (ISIC) [40] que es un proyecto diseñado para facilitar la aplicación de imágenes digitales de la piel para mejorar el diagnóstico de melanomas y así reducir la mortalidad de pacientes por esta enfermedad. Tiene una base de datos de más de 10.000 imágenes dermatoscópicas etiquetadas según su diagnóstico histopatológico y es la colección de imágenes dermatoscópicas de lesiones cutáneas más grande disponible públicamente. Todas las imágenes se encuentran en formato JPG.

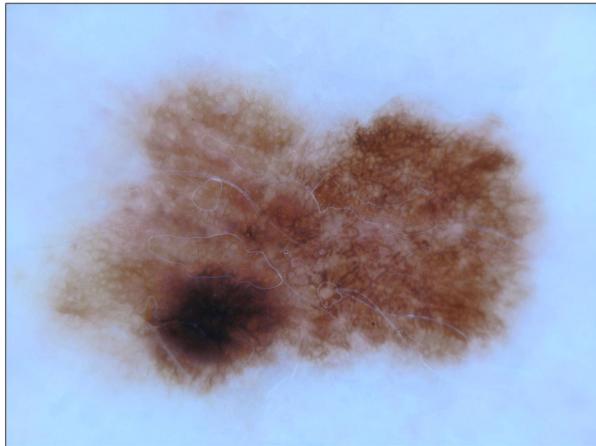
Para la clasificación por métodos de procesamiento de imágenes - y teniendo en cuenta la limitación del tiempo del experto - se eligieron 124 imágenes de esta base de datos, que fueron clasificadas por un dermatólogo experto en dermatoscopia en dos clases: sospechosas y no sospechosas. A su vez el experto identificó la presencia de los patrones de velo azul blanquecino, de red pigmentada atípica y asimetría de estructuras y color, y por último, si las lesiones les resultaban sospechosas, y se utilizó como *Gold Standard* para el cálculo del rendimiento del algoritmo diseñado. Se utilizó el criterio del médico experto como *Gold Standard* por dos razones. En primer lugar, en la base de datos las imágenes no indican si la lesión presenta alguno de los patrones que se analizan en este trabajo, por lo que se requirió que un experto las clasifique manualmente. En segundo lugar, el método de procesamiento de imágenes tiene como objetivo imitar el criterio de un médico experto al realizar un *screening* de un paciente, por lo que el *Gold Standard* del algoritmo debe ser la clasificación que realiza un experto y no los resultados histopatológicos de las lesiones.

Las imágenes fueron seleccionadas manualmente y de manera aleatoria de la base de datos y se revisó que no estuviera incluidas aquellas que presentaran algunas de las siguientes características:

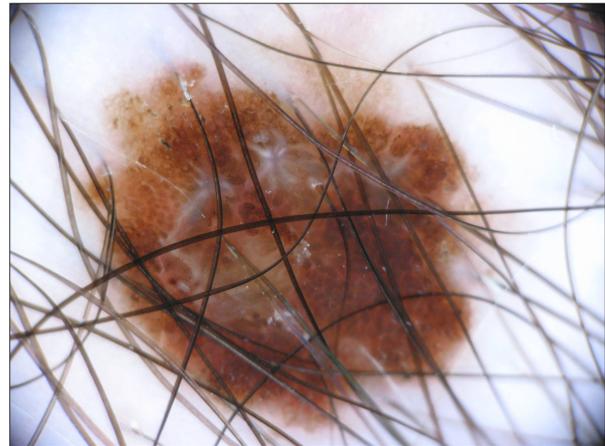
- Imágenes que contienen pelos gruesos: Esto es debido a que los algoritmos utilizados para eliminar el vello grueso deteriora mucho la imagen.
- Imágenes que predomina la luz azul en el fondo: Estas imágenes causan que el algoritmo de detección de velo azul blanquecino presente errores.
- Presencia de artefactos no asociados a la lesión: Algunas imágenes de la base de datos presentan artefactos como marcadores de otros colores o presentan errores en la forma de adquisición, no tratándose de imágenes dermatoscópicas.

- Imágenes con bordes negros: Esto se debe a que estas imágenes dificultan el proceso de segmentación.

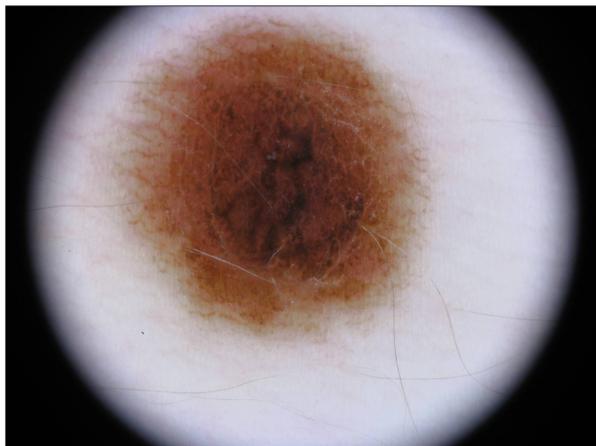
En la figura 9 se muestran cuatro lesiones que representan cada una de las características por las que las imágenes fueron excluidas de la base de datos.



(a) Imágenes con luz azul predominante como luz de fondo



(b) Imágenes con mucha cantidad de pelos gruesos



(c) Imágenes con borde negro



(d) Imágenes no dermatoscópicas

Figura 9: Imágenes excluidas en la base de datos.

En el caso de la red neuronal se utilizaron 4000 imágenes para el conjunto de entrenamiento, 200 imágenes para el conjunto de validación y 604 imágenes para el conjunto de evaluación. Estas imágenes fueron seleccionadas manualmente al azar de la base de datos *ISIC*.

El cuadro 3 se muestra la cantidad de imágenes utilizadas por clase y por conjunto en el método de inteligencia artificial.

En la figura 10 se presentan cuatro ejemplos de imágenes de la base de datos utilizada para el entrenamiento de la red neuronal.

	No Melanoma	Melanoma	Total
Conjunto de entrenamiento	2000	2000	4000
Conjunto de validación	100	100	200
Conjunto de evaluación	307	297	604

Cuadro 3: Distribución de imágenes de cada clase para cada conjunto utilizadas para la red neuronal.

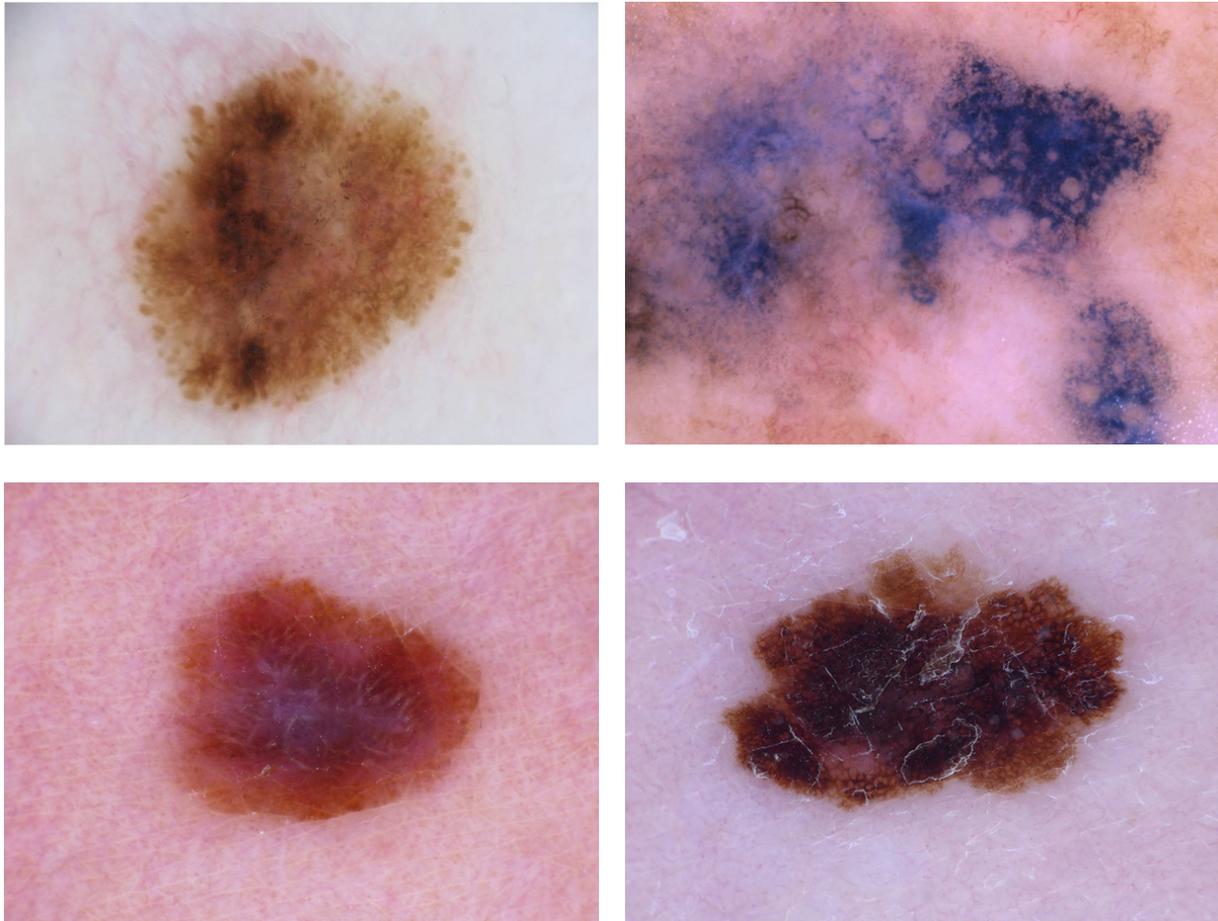


Figura 10: Ejemplo de imágenes utilizadas para el entrenamiento de la red neuronal.

4.3. Método de procesamiento de imágenes

El método propuesto consiste en un algoritmo de tres etapas: el preprocesamiento de la imagen, la segmentación de la lesión y la detección de los patrones y la clasificación del lunar.

La etapa de preprocesamiento consiste en la eliminación de los artefactos presentes en las imágenes, principalmente la presencia de pelos y de ruido general.

Luego, en la etapa de segmentación, se separa la lesión del fondo de la imagen. Una vez segmentada la lesión, se procesa la imagen para la detección de los tres patrones mencionados previamente.

A partir de los patrones detectados, el programa clasificará a la lesión en sospechosa

o no sospechosa.

4.3.1. Preprocesamiento de las imágenes

En todo procesamiento de imágenes se comienza con el acondicionamiento de la imagen. La presencia de pelos y de ruido en las imágenes impide que las posteriores etapas sean correctamente realizadas. En el método desarrollado, la eliminación de pelos y la eliminación de ruido es realizado solamente para la segmentación siguiente. Los métodos de detección de patrones se realizaron sobre las imágenes originales debido a que el preprocesamiento degrada la resolución de la lesión, bajando el rendimiento del mismo.

Se utilizó un algoritmo similar al detector de pelos *Dullrazor* diseñado por Lee T. et al [41]. para la segmentación de los pelos. Las imágenes normalmente poseen el espacio de color RGB, cuyos valores de intensidad se encuentran entre 0 a 255 para cada una de las tres componentes de color. A cada espacio de color se le aplicó la técnica de cerrado de morfología matemática [42] eliminando gran parte de los pelos presentes en la imagen. Cada imagen resultante fue sumada obteniendo nuevamente una imagen RGB. Luego se normalizaron los valores de cada espacio de color dividiendo por 255, llevando las intensidades de cada píxel a un valor entre 0 y 1. Se utilizó un filtro pasa bajos Wiener [43] de tamaño para suavizar la imagen y disminuir los artefactos presentes en la imagen.

En segunda instancia, se utilizó el algoritmo de detección de bordes Canny Edge Detection [44] para encontrar los pelos remanentes en la imagen. Este método utiliza el concepto de histéresis la cual establece dos umbrales que sirven para filtrar los bordes detectados que no pertenezcan a un pelo. Los umbrales establecidos fueron 50 y 70, los cuales son valores de intensidad entre las que se encuentra la intensidad del pelo.

Una vez detectados los pelos, se procedió a eliminarlos de la imagen. En primer lugar se restaron las imágenes de la lesión y la imagen de pelos, resultando en una imagen del lunar sin los pelos detectados. Debido a la resta, los puntos pertenecientes a los pelos adquirieron un valor de intensidad 0. Por cada uno de estos puntos se promediaron las intensidades de los puntos pertenecientes a la vecindad (un *kernel* de tamaño 20x20 píxeles con centro al píxel analizado). Durante el proceso del promedio de las intensidades no se tuvieron en cuenta aquellos puntos con intensidad 0. El valor promedio resultante fue asignado al píxel analizado de cada iteración.

Como se observa en la figura 11 la resolución de la imagen resultante disminuyó considerablemente, reduciendo el rendimiento de los algoritmos de detección de patrones. Es por eso que la imagen es preprocesada sólo para la posterior segmentación. Para la detección de los patrones se aplica a las imágenes de las lesiones con las imágenes originales, con la segmentación realizada.

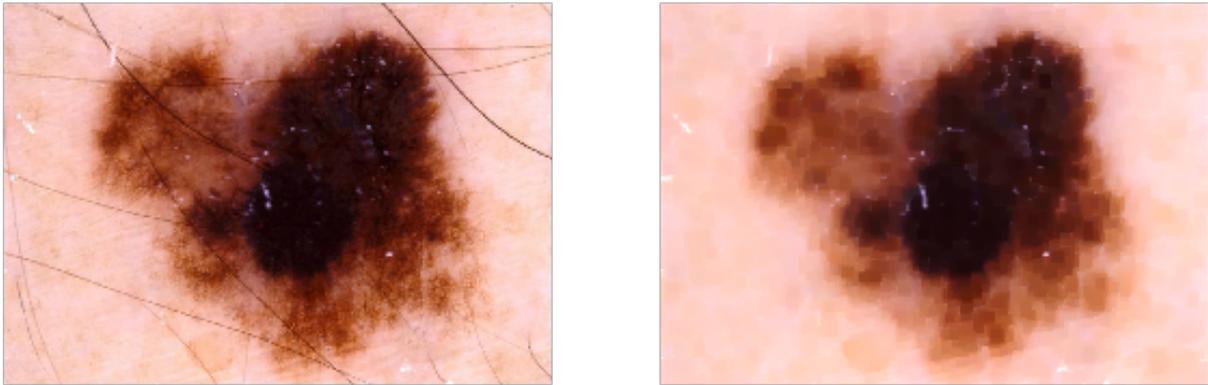


Figura 11: Ejemplo de imagen preprocesada. Se observa que los pelos de la imagen fueron eliminados pero la calidad de la misma fue deteriorada.

4.3.2. Segmentación

Luego del preprocesamiento de la imagen, se segmentó la lesión de la imagen. Para ello se utilizó el algoritmo de K medias, para una segmentación por *clusters* de colores.

El algoritmo de K medias [45] es un método de segmentación automática por *clustering* que permite clasificar un conjunto de elementos según sus propiedades de manera iterativa. Para procesar los datos, el algoritmo comienza seleccionando K centroides al azar, que se utilizan como puntos de inicialización (puntos semillas) en cada grupo. Luego se itera asignando a cada objeto al *cluster* con el punto semilla más cercano. Cuando se terminan de asignar todos los objetos a los *clusters*, se calcula la media aritmética de cada *cluster* y se asignan nuevos puntos semilla. Se recalculan las distancias de cada objeto a los nuevos puntos semilla y se reasignan según su distancia. El algoritmo se repite hasta que los centros de los *clusters* dejen de cambiar. Como resultado se obtiene datos segmentados en *clusters* según sus características. En el método desarrollado, se seleccionaron dos centroides para poder diferenciar el fondo y la lesión.

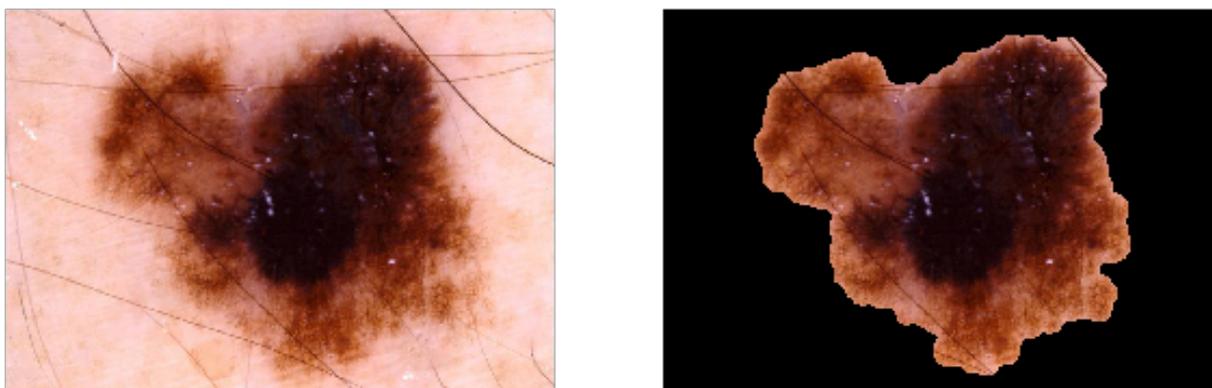


Figura 12: Ejemplo de una segmentación realizada sobre una lesión. Se observa a la izquierda la imagen original y a la derecha la filtrada, donde se pueden distinguir algunos pelos que no fueron filtrados debido a la baja resolución, resultado del paso de preprocesamiento.

Seguidamente, se mejoró la segmentación utilizando técnicas de morfología matemática, puntualmente, utilizando los métodos de cerrado y apertura. Se definió como elemento estructurante un elipse de tamaño 30x30 píxeles, por ser la forma más común de las lesiones. Por último se aplicó dilatación a la imagen para mantener un margen de la lesión a analizar, por posibles patrones presentes en los límites de la lesión.

4.3.3. Detección de patrones

4.3.3.1. Detección de velo azul blanquecino

En las imágenes dermatoscópicas, para identificar el patrón de velo azulado se utilizó el método de K medias que extrae los principales colores de la lesión, anteriormente explicado. En nuestro caso, los *clusters* corresponden a los colores dominantes de la imagen.

Luego de extraer los colores dominantes de la imagen en el espacio de color RGB, se itera sobre los mismos para buscar las diferencias entre un color en particular (en este patrón se utiliza el azul) y los colores dominantes de la imagen. Si la diferencia es menor a un cierto umbral, la imagen se selecciona como correspondiente con el color azul. Es necesario calcular la diferencia y compararlo con un umbral ya que para cada color existen muchos matices. El valor del umbral se debe elegir con precaución pues, según la selección, el algoritmo será más o menos tolerante con respecto a la diferencia entre los colores de imágenes y el color que se desea buscar.

Por último, para detectar el velo azul blanquecino en las imágenes dermatoscópicas, se utilizó una función que itera sobre todas las imágenes y, utilizando los métodos explicados anteriormente, filtra las mismas basadas en su color. Al finalizar el algoritmo, se obtiene un conjunto de imágenes filtradas que tienen el azul como color dominante, correspondientes al patrón de velo azul blanquecino. [46]

4.3.3.2. Detección de red de pigmento atípica

Para la detección de red atípica en las imágenes se utilizó un método semejante al propuesto por Maryam Sadeghi et al. [47] Este método consiste en una etapa de acondicionamiento de la imagen, detección de bordes, selección de estructuras cerradas y por último el criterio de selección. En primer lugar, se aplicó a la imagen un filtro pasa altos para resaltar los detalles de la imagen. Este método fue diseñado por Andrea Polesel et al. [48] para mejoramiento de las imágenes. La aplicación de este último método busca aumentar la nitidez de la lesión y facilitar la distinción de las redes presentes. Los parámetros establecidos fueron $radio = 1$ y $amount = 0.8$.

A continuación, se procedió a detectar los bordes de la imagen para localizar el patrón de redes en la lesión. Para esto se utilizó el detector de bordes *Canny* del paquete de procesamiento de imágenes *Skimage* [49], la cual internamente realiza un suavizado con

un filtro Gaussiano con un parámetro sigma definido (desvío estándar del filtro), una detección de bordes con un filtro *Sobel* [50] y un criterio de histéresis, tal como se describió anteriormente. Para esta etapa de detección se estableció un valor de $\sigma = 2$ y un umbral mínimo, $low_threshold = 0.02$. El resultado de este proceso se puede observar en la figura 13.

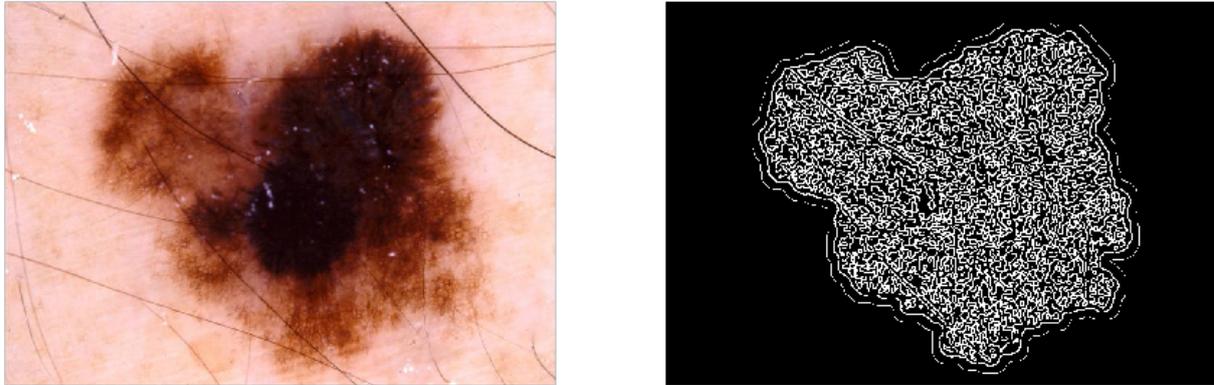


Figura 13: Ejemplo de la operación del filtro pasa altos *Sobel* en la detección de bordes. Se observa que no todos los bordes hallados pertenecen a un borde real de la lesión. Es por eso que la imagen es previamente suavizada con un filtro *Gaussiano* y luego los bordes son filtrados.

A diferencia del modelo propuesto por Maryam Sadeghi et al. [47], no se utilizó la teoría de grafos para diseñar el algoritmo de detección de estructuras cerradas, debido a la ausencia de herramientas en *Python* para el análisis de grafos. El algoritmo de detección diseñado consiste en un proceso iterativo que recorre todos los puntos vecinos encendidos (región 8-conexa) hasta no haber píxeles vecinos disponibles. Éste detectará una estructura cerrada luego de ocho iteraciones en el caso de que el píxel inicial sea nuevamente etiquetado (se contempla la posibilidad de que el píxel inicial tenga todos los píxeles vecinos encendidos). A su vez, se consideró que estructuras mayores a 2500 píxeles y menores a 7 píxeles no son estructuras cerradas.

Por último, se definió un criterio de selección, el cual está compuesto por tres factores: tamaño de los glóbulos, intensidad media del interior del glóbulo en comparación a la intensidad del borde y cantidad de glóbulos dentro de un radio determinado.

En primer lugar, del mismo modo que se hizo en la detección de estructuras, en estos casos se definen límites de mínimo y máximo tamaño que la red puede adquirir. Se comprobó que en varias iteraciones las estructuras con una cantidad de píxeles mayor a 1000 y menor a 50 no pertenecen a una red.

En segundo lugar, se realizó una comparación entre la intensidad media del área encerrada por la red y la intensidad media de los bordes. Según los dermatoscopistas, una de las características de la definición de red es que la intensidad promedio del interior es menor al promedio de los bordes. De esta manera el segundo factor se define a partir de la comparación del promedio de las intensidades en el interior contra el promedio de

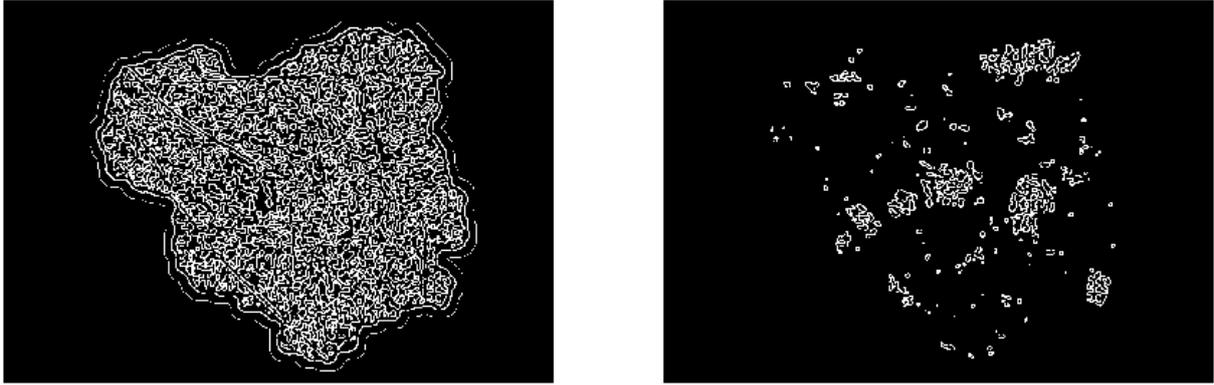


Figura 14: Las bordes detectados son filtrados mediante el algoritmo de detección de estructuras cerradas y de un tamaño específico. Se observa en la imagen de la izquierda, la detección de bordes de la lesión; y en la imagen de la derecha, el filtrado de las estructuras cerradas.

los bordes. Si la intensidad promedio del exterior es mayor al promedio del interior se concluye que la estructura cerrada pertenece a una red atípica.

En tercer lugar, se analizó la cantidad de estructuras cerradas a una distancia determinada. Para ello se calculó para cada estructura su centroide correspondiente, para luego contar la cantidad de centroides dentro de un radio de 50 píxeles. Si la cantidad superaba un valor de diez centroides, se consideró que la imagen poseía un patrón de red atípica.

4.3.3.3. Detección de asimetría

Para el análisis de asimetría de las estructuras y el color de la lesión se diseñó un algoritmo semejante al propuesto por Seidenari et al. [33]. Este método se basa en la partición de la imagen en una cantidad N de cuadros dependiente del tamaño de la imagen y el análisis de las particiones a partir de algunos criterios definidos. En el algoritmo diseñado, el criterio analizado fue el valor medio de intensidad de la imagen total.

El primer paso de este algoritmo consiste en el cálculo de la intensidad media de cada partición de la imagen. Este cálculo es realizado para los tres espacios de color RGB, y son comparados con la media de la imagen total de cada espacio de color, respectivamente. Las particiones fueron etiquetadas si el valor de intensidad fue mayor al criterio establecido.

Luego se dividió la imagen en cuatro cuadrantes y se analizó la cantidad de particiones etiquetados en cada cuadrante. El algoritmo concluye que la imagen posee asimetría si la cantidad de puntos encendidos de uno de los cuadrantes es mayor que la media de las cantidades en los cuatro cuadrantes más un valor arbitrario (para este trabajo se eligió un valor de 3).

4.3.4. Criterio de clasificación

Como se desarrolló anteriormente en este trabajo, el método de los 3 puntos es un método que busca ser altamente sensible. Es por eso que se estableció que el algoritmo clasifique la imagen como sospechosa si al menos uno de los patrones es detectado. De esta manera, aunque la cantidad de falsos positivos sea elevada, se espera que la cantidad de falsos negativos disminuya.

4.4. Método de inteligencia artificial

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son una categoría de las redes neuronales, particularmente efectivas para el reconocimiento de las imágenes y su clasificación. [51]

Se realizan cuatro operaciones principales en este tipo de redes:

1. Convolución
2. No linealidad (*ReLU*)
3. *Pooling*
4. Clasificación (*Fully Connected Layer*)

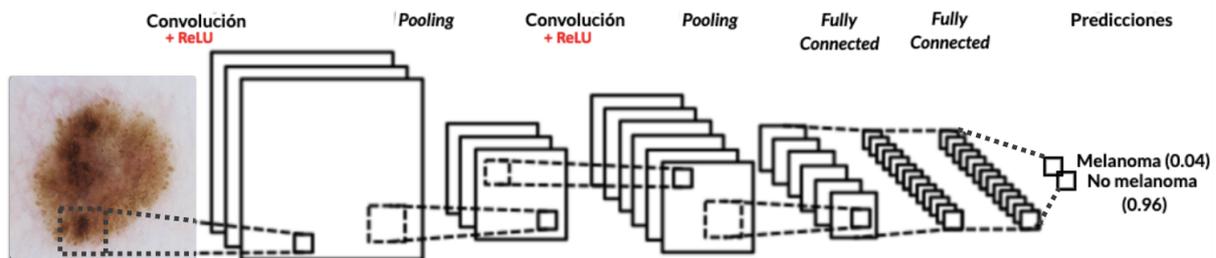


Figura 15: Imagen esquemática de una red neuronal convolucional y sus operaciones. [52]

Las CNN son un método de aprendizaje supervisado y se entrenan utilizando imágenes etiquetadas con sus respectivas clases. Estas redes aprenden la relación entre los objetos de entrada y las etiquetas de clase y utilizan dos componentes: las capas ocultas para extraer las características de cada imagen y, al final del preprocesamiento, las capas *fully connected* para la clasificación. A diferencia de las redes neuronales convencionales, las capas ocultas de una CNN tienen una arquitectura específica, donde no todas las neuronas de una capa están conectadas con todas las neuronas de la siguiente capa, sino con un pequeño grupo, para hacer más eficiente el proceso de aprendizaje. [53]

Los modelos de aprendizaje profundo basados en CNN pueden aprender *hierarchical feature representations*. Esto significa que las características calculadas en las primeras

capas son generales (reconocen formas básicas y pequeñas partes de distintos objetos) mientras que las de las últimas capas son específicas de las imágenes con la que fue entrenada la red. Esta propiedad es la que da lugar para poder utilizar *transfer learning*, una técnica explicada más adelante.

La figura 16 muestra el esquema de una red neuronal convolucional simplificada.

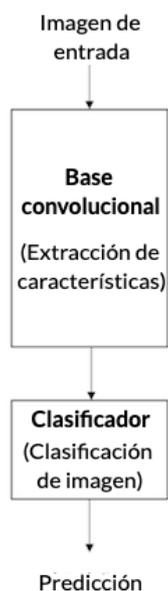


Figura 16: Imagen esquemática del entrenamiento de una red neuronal convolucional y sus operaciones. [54]

4.4.1. Método de redes neuronales

El método propuesto fue implementado en Python [15] utilizando las bibliotecas de TensorFlow [55] y Keras [56] principalmente. TensorFlow [55] es una librería de código abierto para computación numérica y aprendizaje automático a gran escala y Keras es una librería de Python [15] muy utilizada que permite crear modelos de aprendizaje profundo fácilmente.

La arquitectura de la red y los hiperparámetros fueron elegidos teniendo en cuenta la base de datos seleccionada.

4.4.2. Preprocesamiento

Las imágenes dermatoscópicas originales fueron preprocesadas antes de entrenar las redes neuronales convolucionales. Para analizar las imágenes se dividieron en dos clases (melanomas y no melanomas) según la etiqueta del archivo *ISIC*. Para reducir la complejidad del problema, se redimensionaron todas las imágenes a 224x224 píxeles y se normalizaron los valores de los píxeles de las imágenes.

4.4.2.1. *Data Augmentation*

Uno de los problemas más comunes a la hora de entrenar redes neuronales es el *overfitting*. Esto sucede cuando la red no es capaz de generalizar y aprende los casos particulares con los que se entrena, por lo que no tendrá una buena precisión al clasificar nuevas imágenes que no se le han presentado antes. Hay diferentes soluciones para evitar el *overfitting* en modelos de aprendizaje profundo (como la utilización de *Dropouts*, *Transfer Learning*, *Batch Normalization*, entre otros). El *Data Augmentation* tiene un enfoque distinto a los mencionados anteriormente ya que interviene en la base de datos de entrenamiento. Esta técnica crea imágenes modificadas a partir de la base de datos, realizando transformaciones aleatorias tales como rotaciones o reflejos de las imágenes originales, acercamientos, alejamientos, desplazamientos, aplicando distorsiones o cambiando la paleta de colores, entre otras. Debido a que las imágenes disponibles para este trabajo eran limitadas, se le aplicaron transformaciones a las imágenes originales para obtener una mayor cantidad en cada categoría y para que el modelo pueda entrenarse con variaciones en las imágenes para aumentar la precisión a la hora de predecir y evitar el *overfitting*.

Esta técnica se implementó utilizando la librería de *Keras* [56] en *Python 3.6* [15]. Las operaciones que se utilizaron junto con los valores elegidos se muestran a continuación.

```
ImageDataGenerator(rescale = 1./255,
                   rotation_range = 40,
                   width_shift_range = 0.2,
                   height_shift_range = 0.2,
                   shear_range = 0.2,
                   zoom_range = 0.2,
                   horizontal_flip = True,
                   vertical_flip = True,
                   fill_mode = 'nearest')
```

A continuación se describe el funcionamiento de cada una de las operaciones utilizadas. Se agregaron marcos negros a las imágenes para facilitar la visualización de las transformaciones sobre ellas. La figura 17 muestra una imagen original de la base de datos a la que se le aplicarán las distintas transformaciones.

1. *Shift*

Esta operación mueve todos los píxeles de la imagen en una dirección. Puede ser tanto horizontal como vertical manteniendo las dimensiones de la imagen. Se eligió un valor de 0.2, que corresponde a una traslación del 20% de la altura o ancho de la imagen. La traslación se realiza para que la red reconozca el objeto en cualquier parte de la imagen, incluso estando parcialmente en los bordes de la misma.

Las figuras 18 y 19 muestran el resultado de aplicar esta transformación horizontal y vertical, respectivamente, sobre la imagen de la figura 17.



Figura 17: Imagen original a la que se le aplicarán las transformaciones.

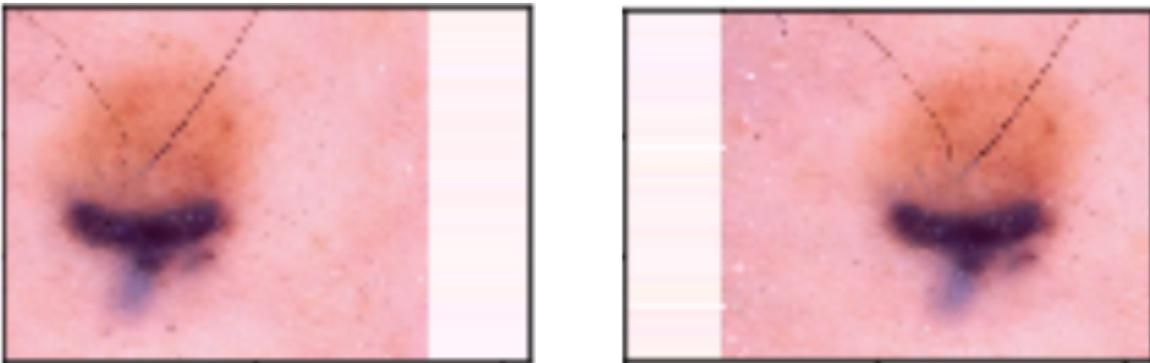


Figura 18: Ejemplo de imágenes luego de la operación de *shift* horizontal.

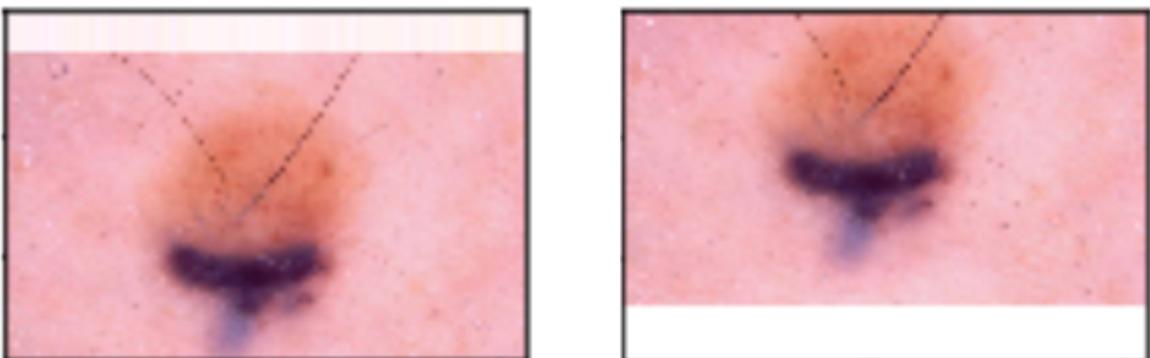


Figura 19: Ejemplo de imágenes luego de la operación de *shift* vertical.

2. Rotación

La rotación gira aleatoriamente la imagen en el sentido de las agujas del reloj en un número determinado de grados (en el rango de 0 a 360) - para este trabajo se eligió 40 -. Este procedimiento deja píxeles fuera del marco de la imagen y esto se realiza para que la red reconozca los objetos presentes en cualquier orientación.

La figura 20 muestra un ejemplo de dos imágenes luego de realizar esta transformación.

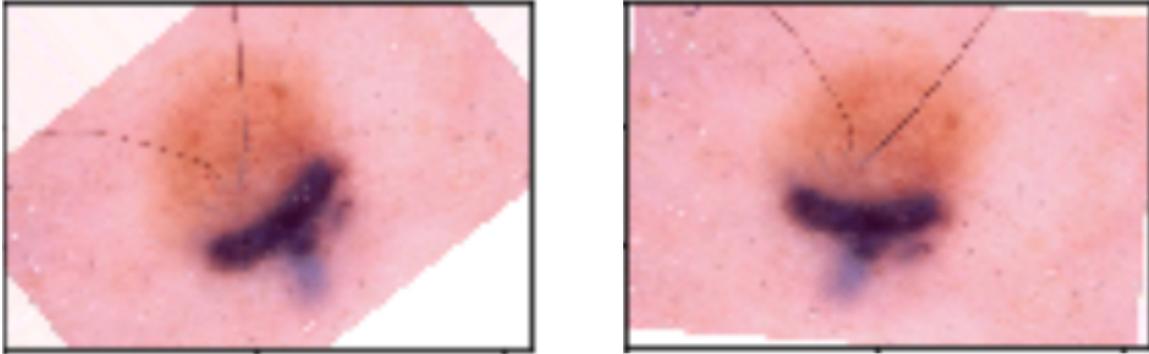


Figura 20: Ejemplo de imágenes luego de la operación de rotación.

3. *Flip*

Esta operación implica invertir las filas o columnas según se trate de un giro vertical u horizontal, respectivamente. Esto se utiliza para que la red elimine el sesgo de suponer que ciertas características de un objeto ocurren en un solo lado en particular.

En las figuras 21 y 22 se observan ejemplos de *flip* vertical y *flip* horizontal, respectivamente.

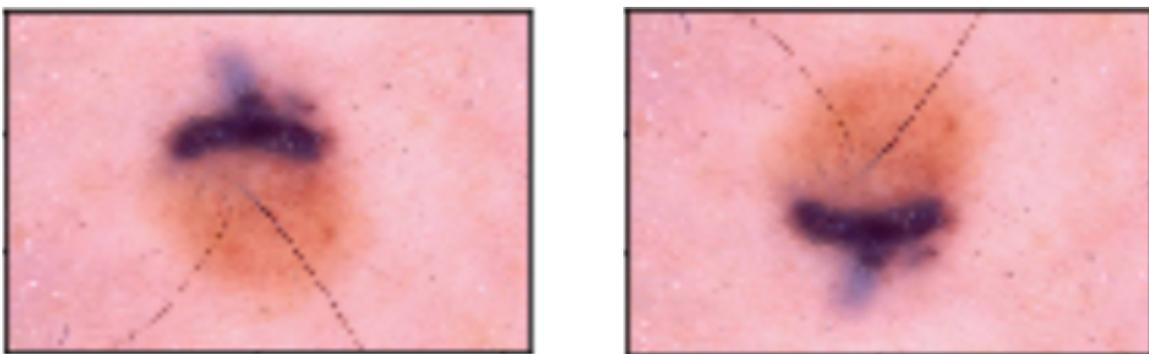


Figura 21: Ejemplo de imágenes luego de la operación de *flip* vertical.

4. *Zoom*

Un aumento de zoom amplía o aleja aleatoriamente la imagen. En nuestro caso configuramos esta operación en 0.2, lo que implica que el rango será $[0.8, 1.2]$, es decir entre 80 % (ampliar) y 130 % (alejar).

La figura 23 muestra un ejemplo de esta transformación.

5. *Shear*

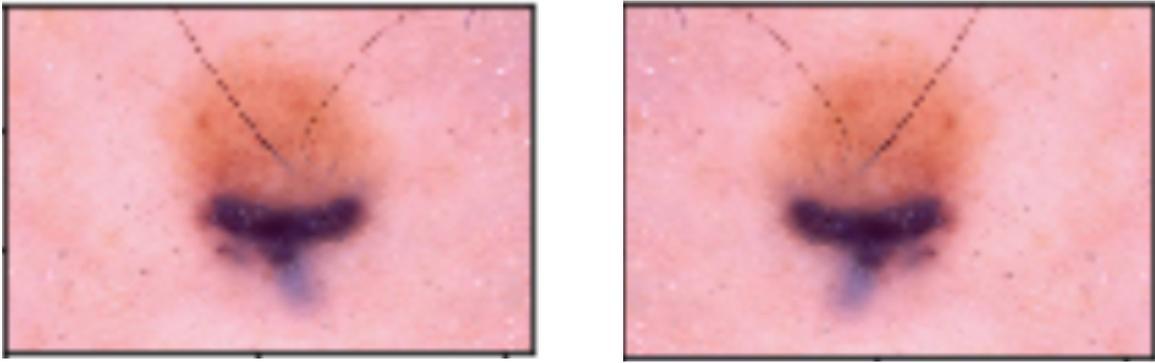


Figura 22: Ejemplo de imágenes luego de la operación de *flip* horizontal.

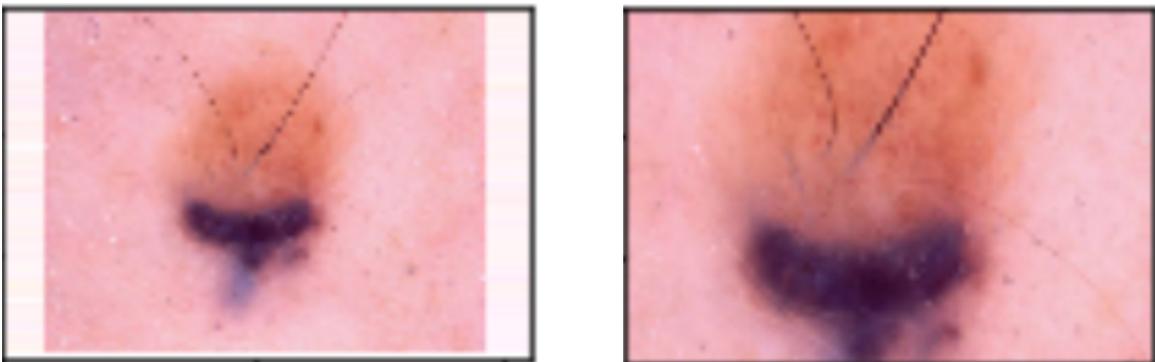


Figura 23: Ejemplo de imágenes luego de la operación de *zoom*.

El *shear* implica distorsionar la imagen en un ángulo en sentido antihorario. Para este parámetro se eligió 0.2 grados.

En la figura 24 se observa un ejemplo del resultado de dos imágenes al aplicarle *shear*.

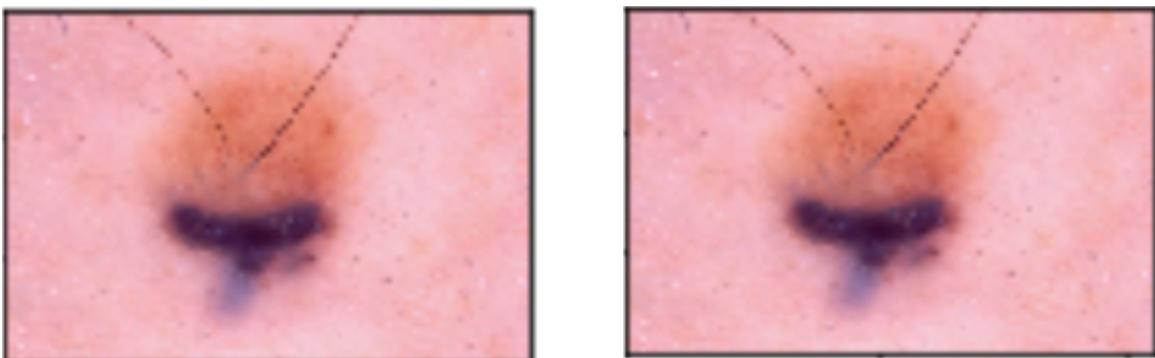


Figura 24: Ejemplo de imágenes al aplicar *shear*.

Luego de implementar *Data Augmentation* a las imágenes originales, se dividieron en forma aleatoria en tres carpetas para utilizarlas en entrenamiento, validación y evaluación.

4.4.3. *Transfer Learning*

Transfer learning es una técnica muy poderosa y utilizada que permite entrenar redes neuronales profundas utilizando considerablemente menos datos que si se las tuviera que entrenar desde un principio, permitiendo así aumentar la precisión de las predicciones y disminuir el tiempo de entrenamiento. Este método consiste en usar pesos de un modelo pre-entrenado y utilizar estas características ya aprendidas para predecir nuevas clases. Estos modelos suelen estar pre-entrenados con millones de imágenes pertenecientes a miles de clases en muchas unidades de procesamiento gráfico (*GPU*) potentes, por muchos días. Para que el modelo funcione bien con un conjunto de datos en particular, se cambia el clasificador del modelo original por un nuevo clasificador y luego se hace *fine tuning* al modelo. Esto implica - en nuestro trabajo - entrenar algunas de las últimas capas del modelo y dejar el resto congeladas. Las capas que están congeladas durante el entrenamiento no cambian sus pesos, mientras que las que están descongeladas se entrenan para mejorar el desempeño de la red en la clasificación de las nuevas imágenes.

El beneficio del *transfer learning* se basa en que no se necesitan enormes cantidades de imágenes en el conjunto de datos de entrenamiento, ni se necesitan grandes recursos a nivel computacional ya que solo se entrenan las últimas capas del modelo.

Algunos ejemplos de arquitecturas que están pre-entrenadas son las ResNet, VGG, Inception, Xception, GoogLeNet, AlexNet, MobileNet, entre otros.

4.4.3.1. **Arquitectura de la red ResNet [57] [58] [59]**

Las redes ResNet [60] surgen de observar que al agregar más capas a las redes neuronales no mejoran la precisión, sino que incluso a veces empeoran, contrario a la creencia de que cuantas más capas profundas, mejora la precisión. Este problema no es debido a *overfitting*, sino debido a un problema de *vanishing gradient*.

Esto ocurre cuando se entrena un modelo de red neuronal muy profundo utilizando técnicas de optimización basadas en gradientes. Ocurre que a medida que se van agregando muchas capas ocultas al modelo, la velocidad de entrenamiento de las capas siguientes aumentan cada vez más rápido. El problema está que cuando se aplica *back propagation* y se calculan los gradientes de *loss* con respecto a los pesos, los gradientes tienden a hacerse cada vez más chicos a medida que se va yendo hacia atrás (*backwards*) en la red. Esto implica que las capas más antiguas aprenden muy despacio comparadas con las capas más nuevas y que los pesos no actualizan sus valores, por lo que no aprenden. Las capas más antiguas son las encargadas de aprender y detectar patrones simples y generales, por lo que si dan resultados poco precisos, el modelo no va a desempeñarse con buenos resultados. En conclusión, el problema de *vanishing gradient* hace que el proceso de entrenamiento tome mucho tiempo y que la precisión de las predicciones del modelo baje.

La arquitectura de la ResNet resuelve este problema haciendo que los gradientes pue-

dan omitir algunas capas (introduce una conexión residual con una capa de identidad). En la figura 25 se muestran las conexiones residuales de la ResNet34 y se compara con otras redes de distinta arquitectura.

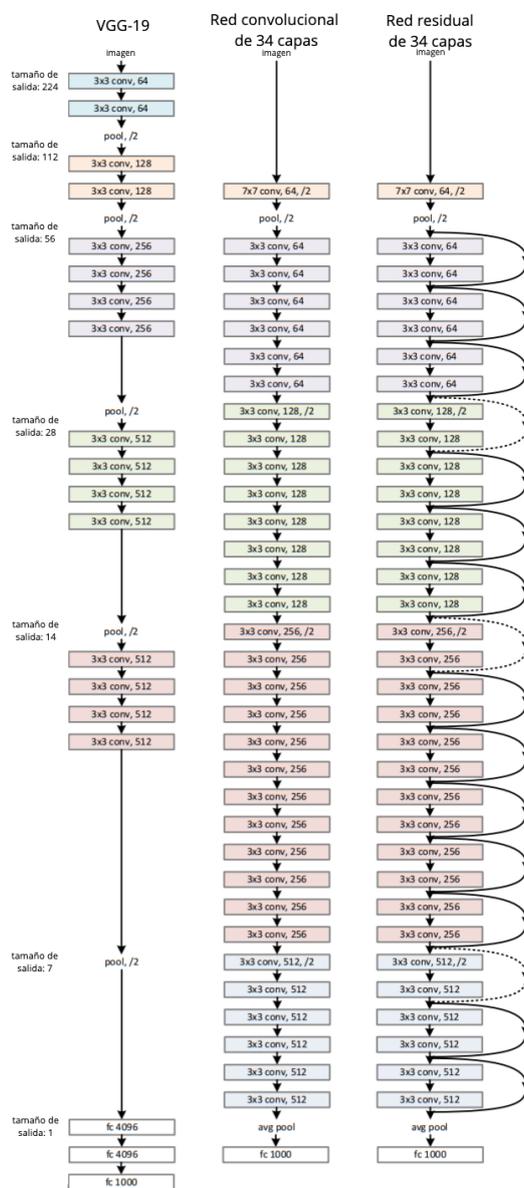


Figura 25: Ejemplos de arquitecturas de redes para ImageNet. Se puede observar el modelo VGG-19 en la figura de la izquierda, una red normal de 34 capas en el centro, y una ResNet de 34 capas a la derecha. [60]

La arquitectura de la red ResNet consiste en un bloque de una convolución y un *pooling*, seguido de cuatro capas de comportamiento similar. Cada una de las capas sigue un mismo patrón. Primero realiza una convolución de 3x3 con un mapa de características de dimensión fija (64, 128, 256, 512, respectivamente), saltando el *input* cada dos convoluciones. Durante la capa entera, las dimensiones del ancho y alto permanecen constantes, aunque se van reduciendo de capa a capa. Esta reducción entre capas se logra aumentando

el *stride* de 1 a 2 en la primera convolución de cada capa, en vez de hacerlo mediante una operación de *pooling*.

En la figura 26 se observa la arquitectura de la ResNet34 donde se pueden identificar los bloques explicados previamente. Los bloques de la imagen representan los volúmenes de las dimensiones de cada capa.

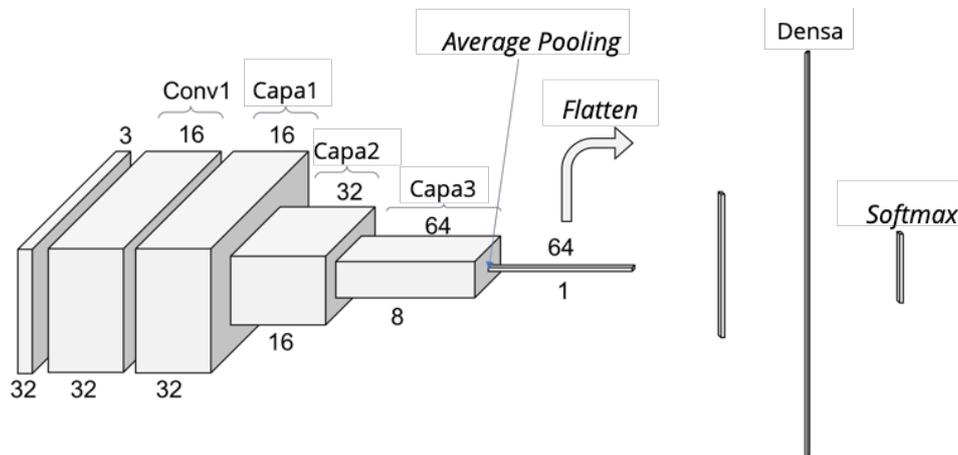


Figura 26: Arquitectura de la red neuronal ResNet34.

Existen distintos tipos de redes ResNet: 18, 34, 50, 101 y 152. Para este trabajo se utilizaron las redes de arquitectura ResNet18 (red residual de 18 capas) y ResNet50 (red residual de 50 capas). La principal diferencia entre ellas no solo radica en la cantidad de capas totales, sino que la profundidad de las capas en cada bloque es distinta: mientras que la ResNet18 tiene dos capas de profundidad en cada bloque, la ResNet50 tiene tres capas por bloque. La figura 27 compara las arquitecturas de redes ResNet con distintas cantidades de capas.

nombre capa	tamaño salida	18-capas	34-capas	50-capas	101-capas	152-capas
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figura 27: Tabla comparativa entre las distintas arquitecturas ResNet.

4.4.3.2. Definición de hiperparámetros

Durante el entrenamiento de las redes se configuraron distintos parámetros para mejorar el desempeño de la red a la hora de clasificar imágenes dermatoscópicas. Los parámetros más importantes que se tuvieron en cuenta son los siguientes:

- *Batch size*: El conjunto de entrenamiento no puede pasar por la red neuronal en una sola pasada debido a las limitaciones de memoria, por lo que se lo divide en conjuntos o números de *batches*. El parámetro del *batch size* define el número de muestras presentes en cada *batch*. Al final de un *batch*, se comparan las predicciones con el valor de salida esperado y se calcula un error. El algoritmo se actualiza a partir de este error y se utiliza para mejorar el modelo.
- *Epoch*: Este parámetro define el número de veces que el algoritmo aprende sobre la base de datos de entrenamiento completa. Una *epoch* es un ciclo completo de aprendizaje.
- *Learning rate*: Define el tamaño del paso con los cuales se actualizan los pesos del modelo con respecto al *gradient descent* estocástico. Como es un hiperparámetro que controla cuánto se ajustan los pesos de la red, tiene mucha importancia al entrenar solamente algunas capas de las redes a diferencia de cuando se deja congelada toda la base convolucional de un modelo. Este parámetro permite asegurarse que los pesos de la red van cambiando al ritmo adecuado.
- *Loss function*: Es una forma matemática de medir la eficacia del algoritmo al modelar la base de datos. Cuanto mejores sean las predicciones, menor será la *loss function*.
- Optimizador: El optimizador vincula la *loss function* y los parámetros del modelo, actualizando los pesos del modelo según la salida de la *loss function*, que marca si el camino tomado es el correcto o no. Todos los optimizadores tienen como objetivo minimizar la *loss function*.

El cuadro 4 muestra los valores de hiperparámetros utilizados para el método de inteligencia artificial propuesto en este trabajo.

<i>Batch Size</i>	<i>Epochs</i>	<i>Learning Rate</i>	<i>Loss Function</i>	Optimizador
20	10 / 50	0.00001	Entropía cruzada binaria	Adam

Cuadro 4: Valores de hiperparámetros elegidos para la red neuronal de arquitectura *Res-Net50*.

4.4.3.3. Nuestra arquitectura

Para este trabajo se partió de la base del modelo ResNet50, inicializándola con los pesos pre-entrenados de ImageNet [61].

Se programó el parámetro `include_top=False` para que no incluya la capa de *pooling* final y la capa de *fully connected* en el modelo original. De esta manera, se eliminó el clasificador del modelo pre-entrenado.

Como nuevo clasificador se agregaron una serie de capas *fully connected* que reciben la información de las características extraídas por la base convolucional. También se agregaron *dropouts* a cada capa *fully connected* para reducir la probabilidad de *overfitting*, ya que esta función ignora aleatoriamente variables de salida de capas ocultas en cada actualización de la fase de entrenamiento, por lo que la red aprende a clasificar imágenes con algunas neuronas reduciendo así el *overfitting* del set de entrenamiento y mejorando la precisión de la red. Por último, se utilizó *Softmax* como función de activación para la última capa. Esta función es una forma de regresión logística que normaliza los valores de entrada en un vector de valores en forma de distribución de probabilidades cuya suma total da 1. La ecuación de esta función diferenciable es

$$\text{softmax}(a_i) = \frac{e^{a_i}}{\sum_{k=1}^K e^{a_k}}$$

Luego se creó un generador de datos para automatizar la obtención de los datos del conjunto de entrenamiento con los siguientes parámetros:

```
ImageDataGenerator(rescale = 1./255,
                   rotation_range = 40,
                   width_shift_range = 0.2,
                   height_shift_range = 0.2,
                   shear_range = 0.2,
                   zoom_range = 0.2,
                   horizontal_flip = True,
                   vertical_flip = True,
                   fill_mode = 'nearest')
```

Para el generador de datos para los conjuntos de validación y evaluación se utilizaron los parámetros:

```
ImageDataGenerator(rescale = 1./255)
```

Con el comando `flow_from_directory` se crearon los generadores de entrenamiento, validación y testeo. Este comando permite crear un flujo continuo para cargar y preparar las imágenes, y genera *batches* de datos normalizados.

Para lograr minimizar la función de costo para obtener el menor valor de error posible

con el fin de aumentar la precisión del modelo, se debe iterar sobre una base de datos de entrenamiento muchas veces, ajustando los parámetros del mismo.

Para los tamaños de los *batches* se definieron los valores:

```
batch_size = 20
```

Luego de programar todos los parámetros para la obtención de los datos, se comenzó congelando la base convolucional del modelo de ResNet50 para aprovechar lo aprendido por el modelo original, y se compiló el modelo congelado junto con el clasificador agregado, utilizando como algoritmo de optimización a *Adam* [62]. Como *loss function* se eligió la entropía cruzada binaria.

Para minimizar el valor del error de la función de costo, se utilizó entropía cruzada binaria [63]. Se eligió el *accuracy* como métrica para evaluar durante el entrenamiento y evaluación del modelo.

Para entrenar el modelo con los datos generado por un generador *batch-by-batch*, se utilizaron los parámetros:

```
model.fit_generator(train_generator ,  
                   epochs = 10 ,  
                   callbacks = checkpoint ,  
                   steps_per_epoch =  
                   number_of_train_images / batch_size ,  
                   validation_data = validation_generator ,  
                   validation_steps =  
                   number_of_val_images / batch_size)
```

Un *callback* es un conjunto de funciones que se aplican en etapas determinadas del proceso de entrenamiento de un modelo para obtener estados internos y estadísticas del mismo durante el entrenamiento. El *checkpoint* es una técnica de tolerancia a fallas para procesos de ejecución largos. Si ocurre un error durante el entrenamiento, se puede utilizar el *checkpoint* (que son los pesos del modelo) como punto de partida para un nuevo entrenamiento, sin perder todo lo aprendido.

Cada paso (*step*, en inglés) del entrenamiento procesa una cantidad del valor de *batch_size* de imágenes que se utilizarán para actualizar los valores del gradiente del modelo. El parámetro de *steps_per_epoch* se calcula dividiendo la cantidad de imágenes del conjunto de entrenamiento por el *batch_size*. Para calcular el parámetro de *validation_steps* se realiza el mismo cálculo pero con la cantidad de imágenes del conjunto de validación.

Luego de la etapa de *fine tuning*, se convirtió el último bloque de la base convolucional de la ResNet50 en entrenable y se compiló el modelo con los mismos parámetros elegidos anteriormente.

Se eligieron valores bajos de *learning rate* (0.00001) ya que si este valor es alto aumenta el riesgo de perder el conocimiento previo de la red pre-entrenada.

Por último, se entrenó el modelo eligiendo los mismos parámetros para el *.fit_generator*, excepto por el número de *epochs* que se cambió a 50.

Luego del entrenamiento largo, se calcularon distintas métricas de evaluación del modelo y se graficaron los valores.

4.5. Métricas de evaluación

Para evaluar la efectividad de las redes neuronales durante el diseño de la arquitectura y el entrenamiento, se tomaron en cuenta los valores de *accuracy*, *validation accuracy*, *total loss* y *validation loss*. Se consideraron las arquitecturas con valores más altos de *accuracy* y con valores más bajos de *loss*.

Para la evaluación cuantitativa del rendimiento de los métodos se utilizaron las siguientes métricas definidas como:

- Precisión = $\frac{VP+VN}{VP+VN+FP+FN}$
- Especificidad = $\frac{VN}{VN+FP}$
- Sensibilidad = $\frac{VP}{VP+FN}$
- Valor predictivo positivo (VPP) = $\frac{VP}{VP+FP}$
- Valor predictivo negativo (VPN) = $\frac{VN}{VN+FN}$

Donde:

- Verdadero positivo (VP): el número de lesiones clasificadas correctamente como sospechosas (Método de Procesamiento de Imágenes) o melanomas (Método de Inteligencia Artificial).
- Verdadero negativo (VN): el número de lesiones clasificadas correctamente como no sospechosas (Método de Procesamiento de Imágenes) o benignas (Método de Inteligencia Artificial).
- Falso positivo (FP): el número de lesiones clasificadas incorrectamente como sospechosas (Método de Procesamiento de Imágenes) o melanomas (Método de Inteligencia Artificial).
- Falso negativo (FN): el número de lesiones clasificadas incorrectamente como no sospechosas (Método de Procesamiento de Imágenes) o benignas (Método de Inteligencia Artificial).

Se utilizó una métrica adicional para la segmentación de las lesiones en el método de procesamiento de imágenes definida como:

$$\blacksquare \text{ Proporción de segmentaciones buenas} = \frac{BS}{BS+SPE+STE}$$

Donde:

- Buenas segmentaciones: Segmentaciones similares a las realizadas por el experto consultor y/o a las máscaras de segmentación de la base de datos *ISIC*.
- Segmentaciones parcialmente erróneas: Segmentaciones que abarcan el 75% de la lesión, o segmentaciones que incluyen hasta un 30% de la imagen no pertenecientes a la lesión.
- Segmentaciones totalmente erróneas: Segmentaciones que incluyen más de un 30% del fondo de la imagen, o que presenta ausencia completa de segmentación.

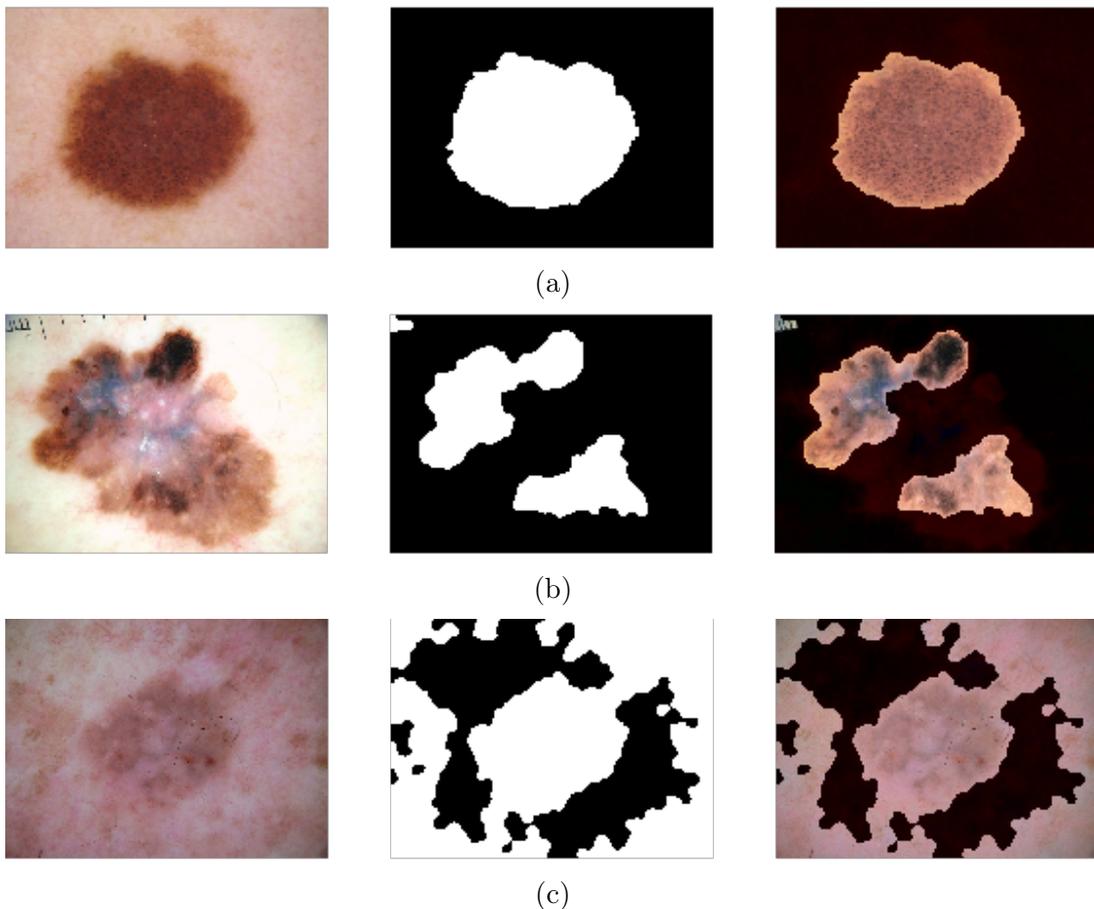


Figura 28: (a) Segmentación buena, (b) Segmentación parcialmente errónea, (c) Segmentación totalmente errónea

5. Resultados

5.1. Método de procesamiento de imágenes

El *Gold Standard* se definió como el juicio del experto, quien analizó la presencia de los tres patrones en cada imagen de las 124 del conjunto de evaluación utilizado y las clasificó como sospechosa o no sospechosa. Luego, se evaluaron estas imágenes en el software propuesto y se compararon los resultados de cada patrón analizado por separado y por último el resultado final de clasificación (sospechoso o no sospechoso), contrastándolo con el *Gold Standard*.

En el cuadro 5 se presenta la matriz de confusión que se utilizará para representar los resultados:

		Predicciones	
		Sospechoso	No sospechoso
Valores reales	Sospechoso	VP	FN
	No sospechoso	FP	VN

Cuadro 5: En donde las siglas son: VN = Verdadero Negativo, VP = Verdadero Positivo, FN = Falso Negativo y FP = Falso Positivo.

5.1.1. Resultados globales

Con el criterio de clasificación establecido explicado previamente, y contrastando esta clasificación con el *Gold Standard*, se obtuvieron los siguientes resultados:

		Predicciones	
		Sospechoso	No Sospechoso
Valores reales	Sospechoso	82	19
	No Sospechoso	15	8

Cuadro 6: Resultados globales

Cuyas métricas estadísticas son:

- Precisión = 72.58 %
- Sensibilidad = 81.20 %
- Especificidad = 34.78 %
- VPP = 84.54 %
- VPN = 29.63 %

5.1.2. Resultados de segmentación

En 44 de las 124 imágenes utilizadas, se presentó algún tipo de error (segmentación parcialmente errónea o totalmente errónea según la métrica de segmentación propuesta) en la segmentación de las lesiones.

En el cuadro 7 se presenta una tabla con la cantidad de imágenes pertenecientes a cada clase de segmentación según su resultado.

Segmentación	Buena	Parcialmente errónea	Totalmente errónea
Casos	80	32	12

Cuadro 7: Cantidad de imágenes por cada tipo de segmentación

El resultado obtenido para la métrica de segmentación es:

- Proporción de segmentaciones buenas = 64.52 %

5.1.3. Resultados de cada uno de los patrones detectados

A continuación se presentarán los resultados obtenidos para la detección de los tres patrones analizados.

5.1.3.1. Detección de velo azul blanquecino

El cuadro 8 muestra la matriz de confusión obtenida para el patrón de velo azul blanquecino.

		Predicciones	
		Sí	No
Valores reales	Sí	25	55
	No	24	20

Cuadro 8: Matriz de confusión de los resultados de la detección de velo azul blanquecino.

A partir de la matriz de confusión, se obtuvieron las siguientes métricas:

- Precisión = 64.52 %
- Sensibilidad = 55.56 %
- Especificidad = 69.62 %
- VPP = 51.02 %
- VPN = 73.33 %

Luego, se implementó el método diseñado y se variaron los umbrales del algoritmo entre los valores 70 y 83.

La figura 29 muestra la curva ROC (Característica Operativa del Receptor) obtenida.

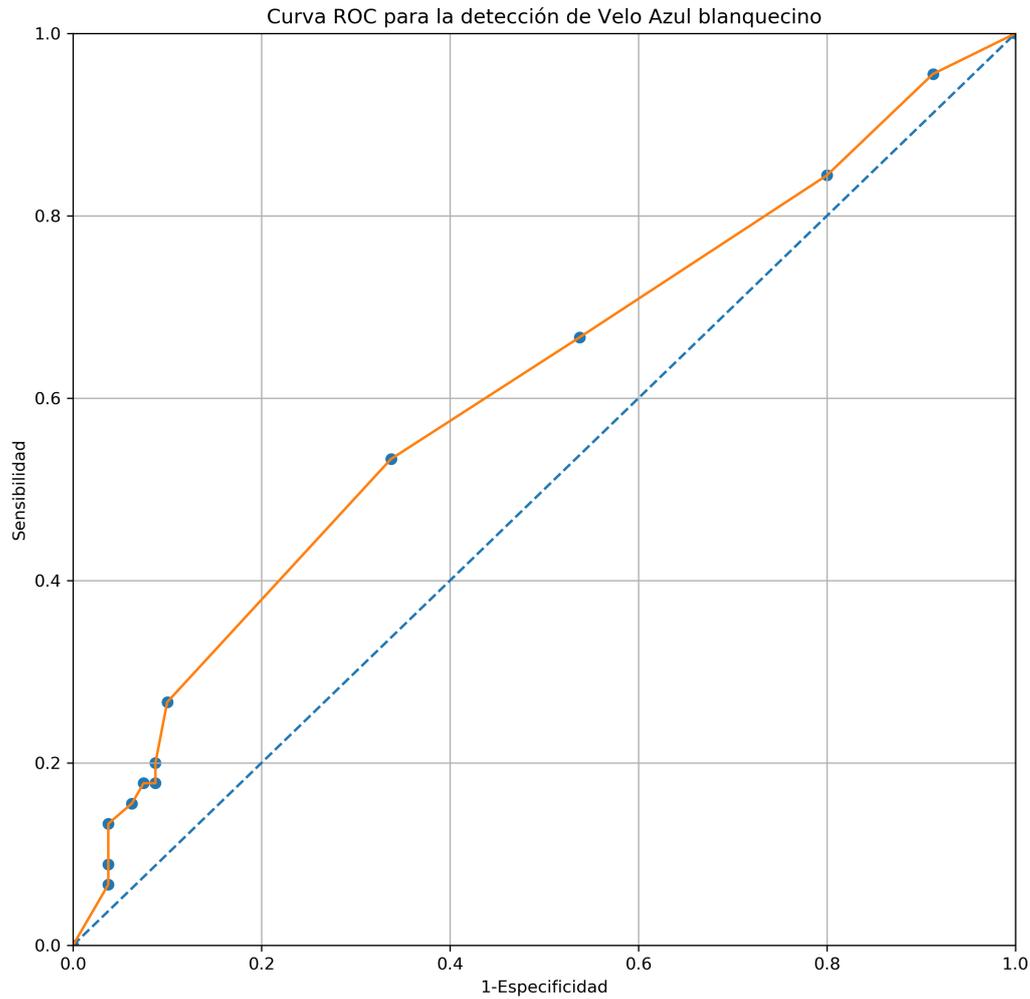


Figura 29: Curva ROC del método de detección de velo azulado.

5.1.3.2. Detección de asimetría

El cuadro 9 muestra la matriz de confusión obtenida para la detección de asimetría en las lesiones.

Asimetría		Predicciones	
		Sí	No
Valores reales	Sí	39	40
	No	12	33

Cuadro 9: Matriz de confusión de los resultados de la detección de asimetría en las lesiones.

A partir de estos resultados, se obtuvieron las siguientes métricas:

- Precisión = 58.06 %
- Sensibilidad = 49.37 %
- Especificidad = 73.33 %
- VPP = 76.47 %
- VPN = 45.21 %

5.1.3.3. Detección de red de pigmento atípica

El cuadro 10 muestra la matriz de confusión obtenida para la detección del patrón de red de pigmento atípica.

Red de pigmento atípica		Predicciones	
		Sí	No
Valores reales	Sí	20	33
	No	33	38

Cuadro 10: Matriz de confusión de los resultados de la detección de red de pigmento atípica.

Las métricas obtenidas para este patrón se muestran a continuación:

- Precisión = 46.77 %
- Sensibilidad = 37.74 %
- Especificidad = 53.52 %
- VPP = 37.74 %
- VPN = 53.52 %

5.2. Método de inteligencia artificial

El método de inteligencia artificial descrito se aplicó al conjunto de evaluación de 604 imágenes dermatoscópicas (297 imágenes de melanomas y 307 de no melanomas) de la base de datos *ISIC*, cuyas imágenes no se utilizaron para entrenar, por lo que resultaron ser nuevas para los modelos. Los entrenamientos de las redes neuronales fueron realizados durante el desarrollo del proyecto.

En el cuadro 11 se presentan las métricas obtenidas al evaluar los resultados del conjunto de entrenamiento en dos modelos de arquitectura ResNet50 y ResNet18.

Los valores obtenidos para las distintas pruebas efectuadas del modelo 1 se muestran en el cuadro 12.

	Precisión (%)	Sensibilidad (%)	Especificidad (%)	VPP (%)	VPN (%)
ResNet50 (Modelo 1)	78.15	89.90	66.78	72.39	87.23
ResNet18 (Modelo 2)	56.62	72.15	54.29	19.19	92.83

Cuadro 11: Comparación de las métricas obtenidas utilizando las arquitecturas ResNet50 y ResNet18 como base convolucional.

	Precisión (%)	Sensibilidad (%)	Especificidad (%)	VPP (%)	VPN (%)
Original (100 %)	78.15	89.90	66.78	72.36	87.23
Prueba 1 (75 %)	71.36	82.15	60.91	67.03	77.92
Prueba 2 (60 %)	73.18	87.54	59.28	67.53	83.11
Prueba 3 (50 %)	72.85	84.51	61.56	68.02	80.43

Cuadro 12: Comparación de las métricas obtenidas utilizando la arquitectura ResNet50, variando la cantidad de imágenes con las que se entrenó a la red. El porcentaje de la primera columna corresponde a la fracción de imágenes utilizadas con respecto al conjunto de entrenamiento completo.

Las figuras 30, 31 y 32 corresponden a la variación de la precisión y sensibilidad según el tamaño del conjunto de imágenes de entrenamiento de ambos modelos.

ResNet50: Precisión vs. Tamaño del conjunto de entrenamiento

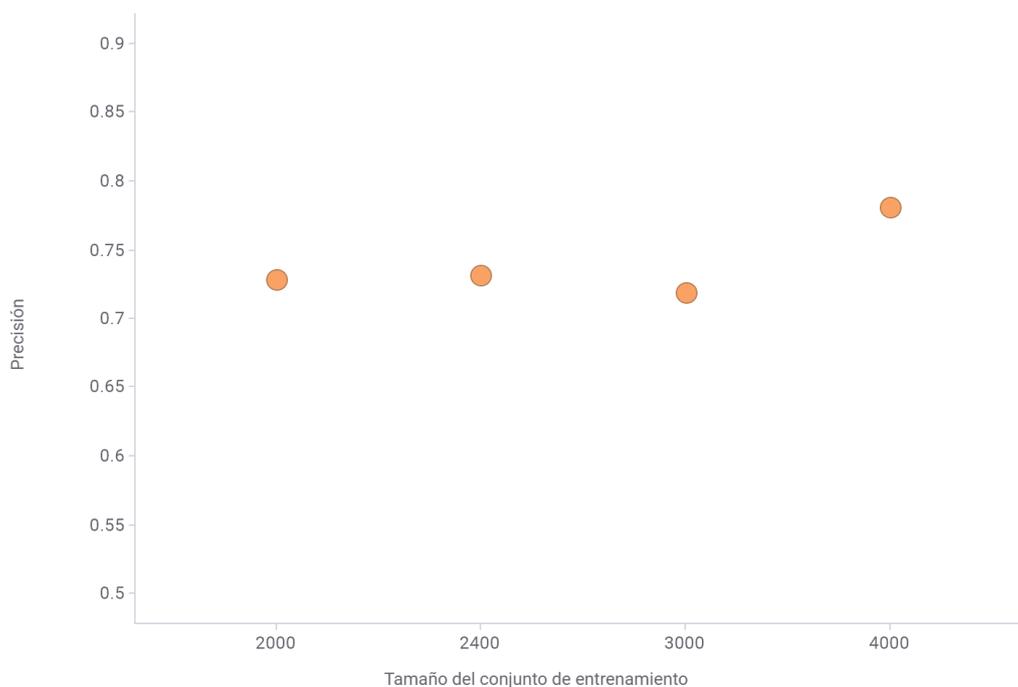


Figura 30: Precisión diagnóstica en función del tamaño del conjunto de imágenes de entrenamiento para la arquitectura ResNet50.

ResNet18: Precisión vs. Tamaño del conjunto de entrenamiento

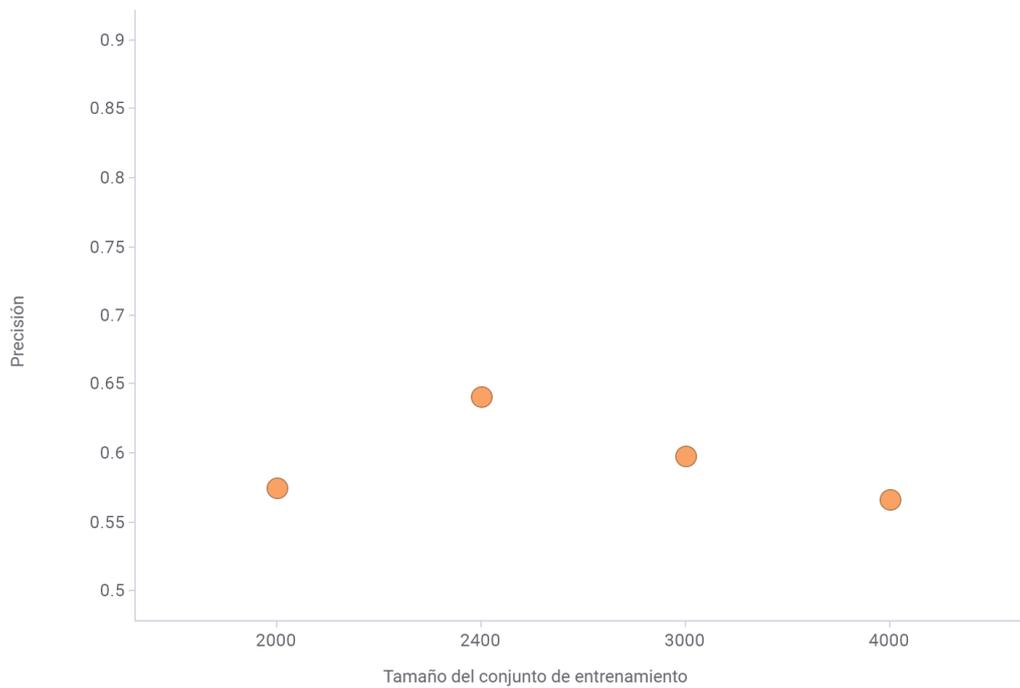


Figura 31: Precisión diagnóstica en función del tamaño del conjunto de imágenes de entrenamiento para la arquitectura ResNet18.

ResNet50: Sensibilidad vs. Tamaño del conjunto de entrenamiento

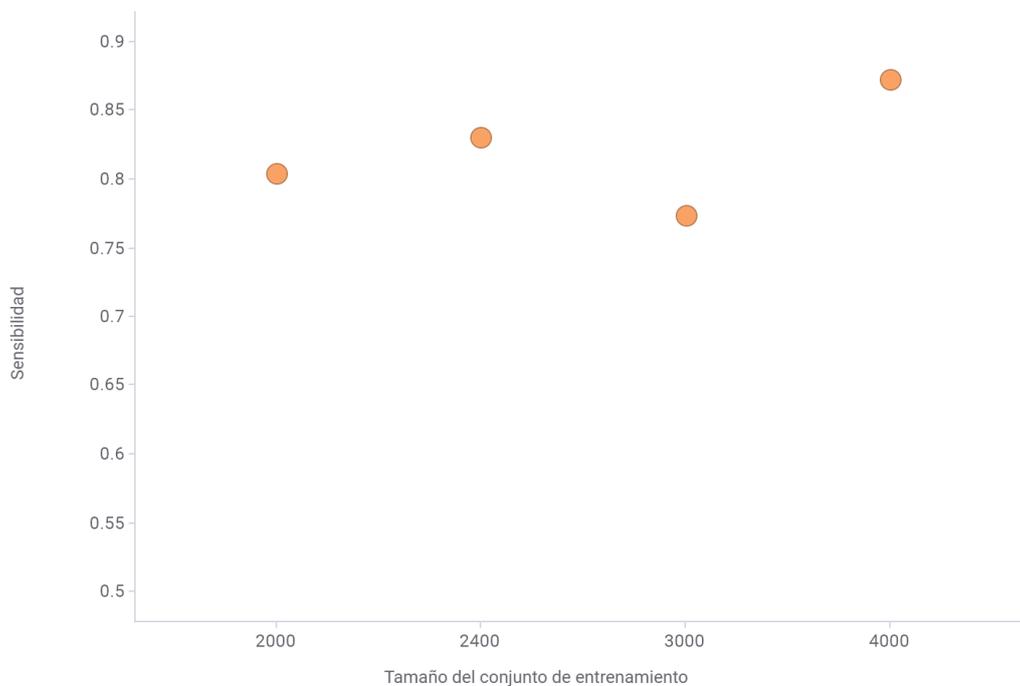


Figura 32: Sensibilidad en función del tamaño del conjunto de imágenes de entrenamiento para la arquitectura ResNet50.

Por último, en los cuadros 13, 14, 15 y 16 se presentan las matrices de confusión obtenidas al evaluar cada prueba del modelo 1.

Original		Predicciones	
		Melanoma	No melanoma
Valores reales	Melanoma	267	30
	No melanoma	102	205

Cuadro 13: Matriz de confusión de la prueba original del modelo con arquitectura ResNet50.

Prueba 1		Predicciones	
		Melanoma	No melanoma
Valores reales	Melanoma	244	53
	No melanoma	120	187

Cuadro 14: Matriz de confusión de la Prueba 1 del modelo con arquitectura ResNet50.

Prueba 2		Predicciones	
		Melanoma	No melanoma
Valores reales	Melanoma	260	37
	No melanoma	125	182

Cuadro 15: Matriz de confusión de la Prueba 2 del modelo con arquitectura ResNet50.

Prueba 3		Predicciones	
		Melanoma	No melanoma
Valores reales	Melanoma	251	46
	No melanoma	118	189

Cuadro 16: Matriz de confusión de la Prueba 3 del modelo con arquitectura ResNet50.

5.3. Comparación de métodos

En el cuadro 17 se observan las métricas obtenidas al comparar los resultados obtenidos por el algoritmo propuesto con el *Gold Standard* y el *Ground Truth*, y aquellos obtenidos al comparar el *Gold Standard* con el *Ground Truth*.

	Precisión	Sensibilidad	Especificidad	VPP	VPN
Método propuesto vs. <i>Gold Standard</i>	72.58 %	81.20 %	34.78 %	84.54 %	29.63 %
Método propuesto vs. <i>Ground Truth</i>	50.81 %	87.5 %	27.63 %	43.30 %	77.78 %
<i>Gold Standard</i> vs. <i>Ground Truth</i>	54.03 %	95.83 %	27.63 %	45.54 %	91.30 %
Red Neuronal vs. <i>Ground Truth</i>	72.58 %	62.50 %	78.95 %	65.22 %	76.92 %

Cuadro 17: Comparación del método diseñado, la red neuronal, el *Gold Standard* y el *Ground Truth*

6. Discusión

Como se puede observar en los resultados obtenidos, existe una diferencia significativa entre el método basado en procesamiento de imágenes y el método basado en inteligencia artificial. El método basado en inteligencia artificial obtuvo un desempeño ampliamente superior, especialmente cuando se evalúa la especificidad. Esto se debe principalmente al método de clasificación elegido para el módulo de procesamiento de imágenes - método de los 3 puntos -, que consiste en un método con énfasis en la sensibilidad. Por otro lado se observa que el rendimiento en la detección de los patrones particulares en este método es baja, causando así un bajo rendimiento general.

6.1. Método de procesamiento de imágenes

En el cuadro 6 se presentan el resultado de la clasificación del método propuesto comparado con la clasificación del experto (*Gold Standard*). Se considera como *Gold Standard* a la clasificación efectuada por un médico especialista, que a diferencia de los resultados histopatológicos de las lesiones, este criterio es subjetivo pues depende de la persona que la realiza. El algoritmo diseñado en este trabajo se basó en el criterio de clasificación del médico experto e intenta imitar los resultados obtenidos por el mismo al realizar *screening* de las lesiones. Esta técnica busca priorizar el reconocimiento de todos los melanomas, por lo que es mucho más sensible que específica.

Al compararlo con el *Gold Standard*, este método presenta una precisión (72.58 %) y una sensibilidad (81.2 %) altas, a diferencia de la especificidad (34.78 %), tal como se espera del método. Aunque los resultados finales fueron buenos, el desempeño de los algoritmos de detección de cada patrón en particular fueron inferiores a los esperados.

Por otra parte, se evaluaron los desempeños del método propuesto y del *Gold Standard* con los resultados histopatológicos de las lesiones (*Ground Truth*). Al observar los resultados se comprueba que la sensibilidad es alta (87.5 % para el método propuesto y 95.83 % para el *Gold Standard*) y la especificidad baja (27.63 % y 27.63 %, respectivamente). Los resultados obtenidos del método diseñado se aproximan bastante a aquellos obtenidos para el *Gold Standard*.

Se observan en los cuadros 8, 9 y 10 los resultados de los algoritmos de detección de velo azul blanquecino, asimetría y red atípica respectivamente. Como se observa en los resultados, el método con mejor desempeño fue el algoritmo de detección de velo azul. Esto se puede deber a que su detección es más simple en comparación a la detección de las demás. A continuación se presentarán posibles causas explicando el bajo desempeño de los algoritmos detectados.

6.1.1. Problemas de segmentación

La principal causa de los errores del algoritmo se debe a la segmentación errónea de las lesiones. Como se presenta en el cuadro 7, aproximadamente un 35 % de las imágenes presentan una segmentación con errores. Las segmentaciones erróneas fueron clasificadas en 2 clases: las parcialmente erróneas y totalmente erróneas.

La causa principal de la segmentación totalmente errónea se debe a la imposibilidad de discriminación de los colores de la lesión con el fondo. La mayoría de las imágenes de esta clase, no presentan bordes nítidamente delimitados.

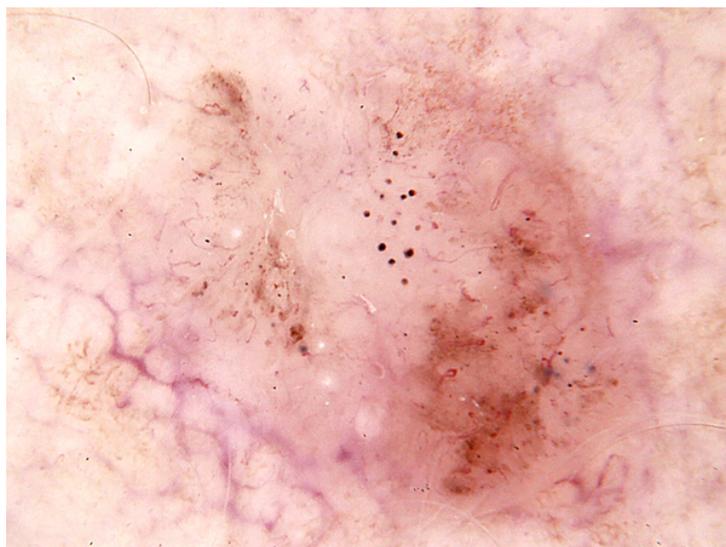


Figura 33: Imagen con segmentación totalmente errónea. Se observa que los colores de la lesión no se pueden diferenciar del fondo de la imagen con claridad. En este caso, el proceso de segmentado fue nulo.

Por otro lado, las imágenes parcialmente erróneas se deben principalmente a dos factores: la lesión presenta cicatrización o iluminación no uniforme. Una lesión presenta cicatrización cuando el sistema inmunológico ataca a un conglomerado cancerígeno resultando en una zona blanca sobre la lesión. Esta diferencia de color, genera un error de discriminación entre la lesión y el fondo de la imagen, por lo que la segmentación no incluye estas regiones. En relación a la iluminación no uniforme, las imágenes que presentan estas características, no pueden ser segmentadas de forma exitosa porque se presentan contrastes de luminancia que alteran los valores de intensidad de la lesión, causando que el algoritmo no puede distinguir los bordes de la lesión correctamente.

6.1.2. Selección del método dermatoscópico a implementar

Como se mencionó anteriormente, el método de los 3 puntos es un método altamente sensible, sin enfoque en la especificidad. Esto se debe a que la cantidad de variables analizadas es reducida, generalizando la clasificación en tres únicos patrones. A continuación

en la figura 34 se presenta un caso donde el médico informó la ausencia de los tres patrones analizados pero aún así se lo consideró como sospechoso.



Figura 34: Lesión sin patrón de velo azul blanquecino, red de pigmento atípica o asimetría.

La lesión presenta otros patrones que la caracterizan como lesión sospechosa. En este caso en particular se presentan en la lesión patrones de puntos, terrones, líneas radiales y pseudópodos, que son patrones dermatoscópicos importantes no contemplados en este método. Es por esto que la implementación de métodos más complejos, como el método de Argenziano [9] o el de Menzies [8], podrían aumentar el rendimiento de la clasificación.

6.1.3. Espacio de color

Las imágenes utilizadas presentan variaciones en color, iluminación y tamaño que aumentan la variabilidad del espacio de trabajo, dificultando la elección de los parámetros utilizados para la clasificación. Aunque todas las imágenes se encuentran en el espacio de color RGB, algunas tienen predominancia de colores particulares, tales como el rojo y el azul. Por ejemplo, las imágenes en las que predomina el color azul causan errores durante la detección de velo azulado, aumentando la cantidad de falsos positivos.

La figura 35 muestra tres ejemplos de imágenes con fondos con distintos colores predominantes.

La utilización de una base de imágenes propias o la adquisición de imágenes bajo un protocolo en común, evitaría la presencia de este error. A su vez, la normalización del espacio de color de las imágenes, como un recurso más del preprocesamiento, podría mejorar el rendimiento del programa para este patrón.

6.1.4. Resolución de las imágenes

La nitidez de las imágenes se presenta como el principal limitante para la detección del patrón de red atípica. La baja resolución de las imágenes se debe principalmente a

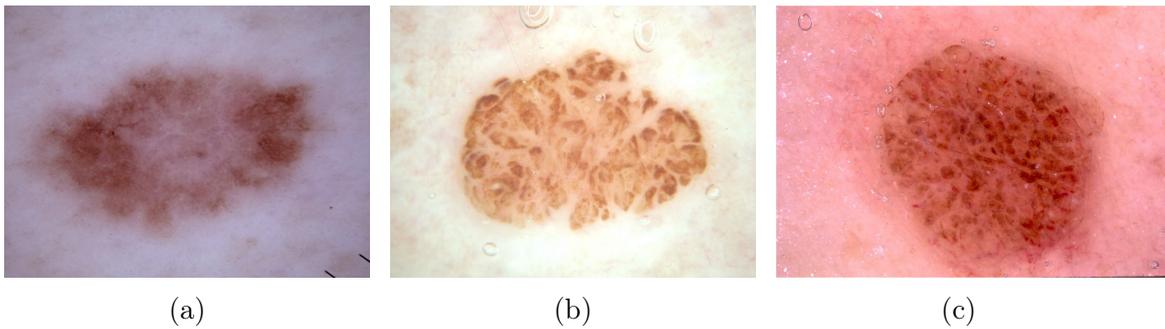


Figura 35: (a) Imagen con fondo con predominancia en azul, (b) Imagen con fondo con predominancia en blanco, (c) Imagen con fondo con predominancia en rojo. Las imágenes poseen diferentes colores predominantes, dificultando la definición de parámetros de clasificación representativos para todas las imágenes.

dos motivos: al origen de las imágenes y al algoritmo de filtrado de pelos. Por un lado, la mayoría de las imágenes obtenidas de la base de datos *ISIC* son un realidad un *zoom* de las imágenes originales, lo que implica una disminución en la resolución de las mismas. También, las imágenes de entrada no poseen un tamaño en común y estas variaciones de tamaños implican una variación de la resolución de cada imagen.

Por otro lado, el proceso de filtrado de pelos presenta una etapa de suavizado (promedios de intensidad), en donde la resolución de la imagen se deteriora. Esta baja resolución de las imágenes dificulta la detección de bordes (es decir, un cambio abrupto de la intensidad del píxel), debido a que se genera una transición suave de intensidades.

6.2. Método de inteligencia artificial

En el cuadro 11 se muestra los rendimientos de las dos arquitecturas de redes neuronales utilizadas para el conjunto de entrenamiento original. Los mejores resultados en todas las métricas - excepto en el valor predictivo negativo - se obtuvieron con la arquitectura de *ResNet50*.

El Modelo 2 tuvo un desempeño muy pobre mientras que el modelo 1 logró una precisión (78.15 %) y una sensibilidad (89.90 %) altas. El motivo por el cual el Modelo 2 tuvo un rendimiento tan bajo se debe a que tuvo una tendencia significativa a predecir las lesiones como si fueran benignas. Esto representa un problema grande teniendo en cuenta que los modelos que se utilizan para predecir diagnósticos médicos, necesitan ser más sensibles que específicos ya que es preferible predecir una lesión benigna como melanoma, antes que predecir una lesión maligna como benigna.

Los resultados obtenidos para el Modelo 1 demuestran que se puede obtener un buen rendimiento al hacer *fine tuning* para clasificar imágenes a partir de un modelo pre-entrenado con la base de datos de *Imagenet*.

Para estudiar el efecto de las variaciones en el número de imágenes del conjunto de entrenamiento, se llevó a cabo una serie de tres pruebas en las cuales se entrenaron y

evaluaron la misma red neuronal (la correspondiente a la arquitectura *ResNet50*), pero con diferentes cantidades de imágenes para el conjunto de entrenamiento de cada prueba.

En el cuadro 12 se pueden observar los resultados obtenidos para cada una de estas pruebas. Como es de esperarse, la precisión de la clasificación baja a medida que la cantidad de imágenes con la que se entrenó la red disminuye. Se puede observar que la Prueba 1 tiene una precisión por debajo de lo esperado. Este número es el resultado del promedio de dos pruebas realizadas, ya que el primer valor que se obtuvo (precisión = 71.36 %) fue menor que el de la Prueba 2 (precisión = 73.18 %). Por este motivo se llevó a cabo otra prueba idéntica donde se cambiaron aleatoriamente las imágenes de la base de datos de entrenamiento y se le realizaron *Data Augmentation* con los mismos parámetros que para el resto de las pruebas.

El resultado de la segunda prueba, aunque mejoró la precisión (precisión = 72.35 %), igualmente tuvo un valor inferior que el de la Prueba 2. Esto se puede deber a que el conjunto de imágenes seleccionadas aleatoriamente del conjunto de entrenamiento original, son las que menos representan las características más importantes de cada clase.

Teniendo en cuenta que la Prueba 1 se realizó dos veces con dos conjuntos de imágenes distintas e igualmente la precisión siguió por debajo de la de la Prueba 2, se puede inferir que el motivo se encuentra en que el conjunto de imágenes con las que se entrenó la Prueba 2 son imágenes particularmente representativas de cada clase y el modelo pudo extraer las características necesarias para clasificar las imágenes del conjunto de evaluación de una manera exitosa.

Observando la figura 30 se puede concluir que el tamaño de la base de datos de entrenamiento y la composición del conjunto de imágenes son variables importantes para la clasificación de imágenes utilizando redes neuronales.

Las matrices de confusión de los cuadros 13, 14, 15 y 16 muestran el desempeño de clasificación de cada prueba realizada en la arquitectura *ResNet50* para el conjunto de imágenes de evaluación. El valor más importante es el que se encuentra en la esquina superior derecha - falsos negativos -, ya que corresponde a los casos de melanoma que fueron clasificados como benignos. Al observar estos valores en cada matriz de confusión, se observa que la prueba que tiene mejor rendimiento es la original que cuenta con 30 casos de falsos negativos en un total de 604 imágenes, reafirmando los resultados discutidos previamente.

6.2.1. Segmentación en el preprocesamiento de las imágenes

Una de las líneas en las que se podrían trabajar a futuro se trata de incluir un paso en el preprocesamiento de las imágenes para segmentar las lesiones. Entrenar las redes neuronales con las lesiones segmentadas podría aumentar la precisión en la clasificación de imágenes y disminuir los tiempos de entrenamiento.

6.2.2. Composición del conjunto de entrenamiento

Foody et al. [64] presentaron una publicación donde demostraron que la composición del conjunto de entrenamiento afecta los resultados de clasificación. En este trabajo se asumió a fines prácticos que cada clase tiene la misma probabilidad de ocurrencia *a priori*. Es por esto que las redes neuronales se entrenaron con la misma cantidad de imágenes para cada clase, sin embargo en los exámenes dermatoscópicos, la cantidad de lesiones benignas supera ampliamente a la cantidad de lesiones malignas. Si esta probabilidad de ocurrencia se supiera *a priori*, se podría entrenar la red neuronal con pesos diferenciados para cada clase, aumentando así la precisión de la misma al clasificar imágenes dermatoscópicas.

6.2.3. Cantidad de imágenes del conjunto de entrenamiento

En este trabajo se utilizaron un total de 2000 imágenes dermatoscópicas de cada clase para entrenar las redes neuronales. También se llevaron a cabo pruebas donde se entrenaron redes neuronales con menor cantidad de imágenes para comprobar la relación entre el desempeño de la red y el tamaño del conjunto de imágenes de entrenamiento. Como era de esperarse, al aumentar el conjunto de entrenamiento con la que se entrena una red, aumenta el rendimiento de la red neuronal. Es por esto que un aumento en la cantidad de imágenes de entrenamiento, debería mejorar la precisión diagnóstica de este modelo.

6.2.4. Considerar los metadatos para el entrenamiento de la red neuronal

Considerar los metadatos del paciente tales como la edad, antecedentes familiares de melanomas, sexo, zona anatómica donde se encuentra el lunar y tipo de piel en el entrenamiento de la red, pueden ayudar a darle un contexto a cada imagen y a que la precisión de la red neuronal aumente. Esto sucede ya que, por ejemplo, si el paciente tiene algún familiar con antecedentes de melanoma, tiene más probabilidades de desarrollar esta enfermedad que otro paciente sin familiares con antecedentes de este tumor.

6.3. Comparación de métodos

Además de comparar el método de procesamiento de imágenes con el Gold Standard y el método de inteligencia artificial con el Ground Truth, se llevó a cabo una comparación entre ambos métodos. Debido a que la evaluación de los métodos por separado se realizó con conjuntos de evaluación diferentes, y fueron comparados contra distintos criterios (criterio del experto vs. resultados histopatológicos, respectivamente), los resultados no son comparables. Por lo tanto, se evaluaron ambos métodos con el mismo conjunto de imágenes, y se obtuvieron los resultados del cuadro 17.

En la primera fila se observan los resultados previamente obtenidos para el método de procesamiento de imágenes propuesto al compararlo con el *Gold Standard*. En la segunda y tercera fila, se muestran los resultados obtenidos al comparar el método propuesto y el *Gold Standard* con el *Ground Truth*, respectivamente. La última fila muestra los resultados conseguidos al evaluar la red neuronal *ResNet50* con las 124 imágenes del conjunto de evaluación, contrastándolos con el *Ground Truth*.

Como era de esperarse, los resultados muestran una mejor precisión para el método de inteligencia artificial y una mejor sensibilidad en el algoritmo de procesamiento de imágenes y para el *Gold Standard*. Esto se debe, como se mencionó anteriormente, a la naturaleza de la estrategia de *screening* y al criterio dermatoscópico elegido (regla de los 3 puntos).

Cuando se compara el método propuesto con el *Gold Standard*, se acarrean errores de clasificación de ambos criterios, lo que resulta en una precisión no tan elevada (72.58 %) de este método. Sin embargo, al comparar ambas clasificaciones con el *Ground Truth*, se obtiene un desempeño similar al *Gold Standard*, y se comprueba así la efectividad del método propuesto.

Debido a la subjetividad de la clasificación realizada por médicos y que en este trabajo se tuvo en cuenta solamente la clasificación de un experto, el *Gold Standard* no es el óptimo. Para obtener un *Gold Standard* representativo del criterio médico, es necesario aumentar la cantidad de expertos involucrados en el proceso de clasificación de estas imágenes.

Por otra parte, es posible ver que la sensibilidad del experto es mucho mayor que el de la red neuronal, si bien su precisión es menor. Es posible aumentar la sensibilidad de la red neuronal variando el umbral de selección de la probabilidad de asignación a cada clase, lo cual traerá también aparejado una baja en la exactitud de clasificación de la red neuronal.

7. Conclusiones

Este trabajo propone dos métodos distintos para la clasificación de lesiones a partir de imágenes dermatoscópicas cutáneas. En cuanto al método de procesamiento de imágenes se logró clasificar las 124 lesiones evaluadas con una precisión de 72.00 % y sensibilidad de 81.20 % al contrastarlo con el *Gold Standard*. En relación a los resultados del método de inteligencia artificial, se alcanzó una precisión de 78.15 % y sensibilidad de 89.90 % al contrastar las 604 imágenes evaluadas con los resultados histopatológicos.

Por otro lado se evaluaron ambos métodos con el conjunto de evaluación del algoritmo de procesamiento de imágenes. Este conjunto de evaluación consiste en 124 imágenes dermatoscópicas, compuesto por 48 melanomas y 76 no melanomas según su resultado histopatológico (*Ground Truth*). Los resultados del método de procesamiento de imágenes, del método de inteligencia artificial y el *Gold Standard* fueron comparados con los del *Ground Truth*, obteniéndose una precisión de 50.81 %, 54.03 % y 72.58 % y una sensibilidad de 87.5 %, 95.83 % y 62.5 %, respectivamente. De esta manera se comprueba que el método de procesamiento de imágenes propuesto tiene un rendimiento similar al del experto al compararse ambos con el *Ground Truth*. Por último, el método propuesto de inteligencia artificial, aunque obtuvo una precisión significativamente mayor al método propuesto y al *Gold Standard*, la sensibilidad tuvo un peor desempeño. Esto se debe a que el método de procesamiento de imágenes propuesto y el *Gold Standard* se diseñaron para utilizarse en *screening*, por lo que son altamente sensibles, mientras que el método de inteligencia artificial fue diseñado para clasificar lesiones según su resultado histopatológico, por lo que se prioriza la precisión.

De esta manera se lograron cumplir los objetivos específicos propuestos para este trabajo.

Las investigaciones y avances en el campo del procesamiento de imágenes biomédicas y de aprendizaje profundo aplicado a la salud, están en constante desarrollo en búsqueda de mejorar la calidad de vida de las personas. Distintos modelos matemáticos para evaluar patrones dermatoscópicos, nuevas arquitecturas de redes neuronales y distintos enfoques en estos ámbitos permitirán desarrollar métodos más eficientes y eficaces para clasificar lesiones cutáneas.

A. Apéndice

Nombre de la imagen	<i>Ground Truth</i>	<i>Gold Standard</i>	Procesamiento de imágenes	Red Neuronal
ISIC_0000001.jpg	0	1	1	0
ISIC_0000008.jpg	0	1	1	1
ISIC_0000013.jpg	1	1	0	1
ISIC_0000020.jpg	0	1	1	1
ISIC_0000026.jpg	1	0	1	0
ISIC_0000029.jpg	1	1	1	1
ISIC_0000038.jpg	0	1	1	0
ISIC_0000049.jpg	1	1	1	1
ISIC_0000055.jpg	0	1	1	0
ISIC_0000060.jpg	0	1	1	0
ISIC_0000064.jpg	0	1	0	0
ISIC_0000080.jpg	0	1	1	1
ISIC_0000085.jpg	0	1	1	0
ISIC_0000089.jpg	0	1	1	1
ISIC_0000105.jpg	0	1	1	0
ISIC_0000142.jpg	1	1	1	1
ISIC_0000143.jpg	1	1	1	0
ISIC_0000150.jpg	1	1	1	1
ISIC_0000161.jpg	1	1	1	1
ISIC_0000166.jpg	1	1	1	1
ISIC_0000173.jpg	1	1	1	1
ISIC_0000189.jpg	0	0	0	0
ISIC_0000199.jpg	0	1	1	0
ISIC_0000221.jpg	0	1	1	1
ISIC_0000224.jpg	0	1	1	1
ISIC_0000236.jpg	0	1	1	0
ISIC_0000266.jpg	0	0	1	0
ISIC_0000288.jpg	1	1	1	0
ISIC_0000377.jpg	0	1	1	0
ISIC_0000408.jpg	0	1	1	0
ISIC_0000423.jpg	0	0	1	0
ISIC_0000436.jpg	0	0	0	0
ISIC_0000476.jpg	0	1	0	0
ISIC_0000516.jpg	1	1	1	1

Nombre de la imagen	<i>Ground Truth</i>	<i>Gold Standard</i>	Procesamiento de imágenes	Red Neuronal
ISIC_0000527.jpg	0	0	0	0
ISIC_0000543.jpg	0	1	0	0
ISIC_0000552.jpg	1	1	1	1
ISIC_0001135.jpg	1	1	1	0
ISIC_0010145.jpg	1	1	0	0
ISIC_0010295.jpg	1	1	1	1
ISIC_0010350.jpg	1	1	0	0
ISIC_0010392.jpg	1	1	1	1
ISIC_0010394.jpg	1	1	1	1
ISIC_0010397.jpg	1	1	0	0
ISIC_0010533.jpg	1	1	0	0
ISIC_0010653.jpg	1	1	1	1
ISIC_0010732.jpg	1	1	1	0
ISIC_0010774.jpg	0	1	0	0
ISIC_0010810.jpg	1	1	0	0
ISIC_0010837.jpg	0	1	1	0
ISIC_0010838.jpg	0	0	1	0
ISIC_0010875.jpg	0	1	1	0
ISIC_0010904.jpg	0	1	1	0
ISIC_0010910.jpg	0	1	1	0
ISIC_0010911.jpg	0	1	0	0
ISIC_0010913.jpg	0	1	1	0
ISIC_0010921.jpg	0	1	0	0
ISIC_0010936.jpg	0	1	1	1
ISIC_0010945.jpg	0	1	0	0
ISIC_0010991.jpg	0	0	1	0
ISIC_0011018.jpg	0	0	0	0
ISIC_0011020.jpg	0	1	1	0
ISIC_0011061.jpg	0	1	0	0
ISIC_0011066.jpg	0	1	1	0
ISIC_0011069.jpg	0	1	1	0
ISIC_0011076.jpg	0	1	1	0
ISIC_0011084.jpg	0	1	1	1
ISIC_0011085.jpg	0	1	1	0
ISIC_0011101.jpg	0	0	1	0
ISIC_0011116.jpg	0	1	1	0

Nombre de la imagen	<i>Ground Truth</i>	<i>Gold Standard</i>	Procesamiento de imágenes	Red Neuronal
ISIC_0011148.jpg	0	1	1	0
ISIC_0011163.jpg	1	1	1	0
ISIC_0011164.jpg	0	1	1	0
ISIC_0011177.jpg	0	0	1	0
ISIC_0011183.jpg	0	0	1	0
ISIC_0011196.jpg	1	1	1	0
ISIC_0011198.jpg	0	0	1	0
ISIC_0011203.jpg	0	0	1	0
ISIC_0011211.jpg	0	1	1	0
ISIC_0011218.jpg	0	1	1	0
ISIC_0011229.jpg	0	0	1	0
ISIC_0011261.jpg	0	1	1	0
ISIC_0011275.jpg	0	1	1	0
ISIC_0011279.jpg	0	1	0	0
ISIC_0011282.jpg	0	1	1	0
ISIC_0011292.jpg	0	1	0	0
ISIC_0011294.jpg	0	1	1	0
ISIC_0011301.jpg	0	0	1	0
ISIC_0025097.jpg	1	1	1	1
ISIC_0025491.jpg	0	1	1	1
ISIC_0026120.jpg	1	1	1	1
ISIC_0026427.jpg	0	1	1	0
ISIC_0026489.jpg	1	1	1	0
ISIC_0026560.jpg	0	1	0	1
ISIC_0026600.jpg	0	0	0	0
ISIC_0026608.jpg	0	1	1	1
ISIC_0026668.jpg	1	1	1	0
ISIC_0026744.jpg	0	1	1	1
ISIC_0026745.jpg	1	1	1	1
ISIC_0026764.jpg	0	0	0	1
ISIC_0026909.jpg	1	1	1	1
ISIC_0026974.jpg	0	0	1	0
ISIC_0026979.jpg	0	0	1	0
ISIC_0027005.jpg	0	1	1	1
ISIC_0027042.jpg	0	1	0	1
ISIC_0027060.jpg	1	1	1	1

Nombre de la imagen	<i>Ground Truth</i>	<i>Gold Standard</i>	Procesamiento de imágenes	Red Neuronal
ISIC_0027087.jpg	0	1	0	1
ISIC_0027123.jpg	0	0	0	0
ISIC_0027126.jpg	1	0	1	1
ISIC_0027155.jpg	1	1	1	1
ISIC_0027167.jpg	0	0	0	0
ISIC_0027277.jpg	1	1	1	1
ISIC_0027838.jpg	1	1	1	1
ISIC_0027964.jpg	1	1	1	1
ISIC_0028003.jpg	1	1	1	0
ISIC_0028326.jpg	1	1	1	0
ISIC_0028584.jpg	1	1	1	1
ISIC_0028642.jpg	1	1	1	1
ISIC_0028898.jpg	1	1	1	0
ISIC_0029909.jpg	1	1	1	1
ISIC_0030060.jpg	1	1	1	1
ISIC_0031034.jpg	1	1	1	0
ISIC_0031386.jpg	1	1	1	1
ISIC_0031401.jpg	1	1	1	1

. Referencias

- [1] Geneser, Annemarie Brüel, Erik Ilsø Christensen, Jørgen Trandum-Jensen, and Klaus Qvortrup. *Geneser Histología*. Panamericana, 4th edition, 2015.
- [2] Giuseppe Argenziano, H Peter Soyer, Sergio Chimenti, Renato Talamini, Rosamaria Corona, Francesco Sera, Michael Binder, Lorenzo Cerroni, Gaetano De Rosa, Gerardo Ferrara, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the internet. *Journal of the American Academy of Dermatology*, 48(5):679–693, 2003.
- [3] American cancer society. <https://www.cancer.org/es/cancer/cancer-de-piel.html>. Accessed: 2019-08-20.
- [4] Skin cancer facts statistics. <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/>. Accessed: 2019-08-20.
- [5] Estadísticas - incidencia. <https://www.argentina.gob.ar/salud/instituto-nacional-del-cancer/estadisticas/incidencia>. Accessed:2019-07-28.
- [6] Key statistics for melanoma skin cancer. <https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html>. Accessed: 2019-07-28.
- [7] Kopf AW. Friedman RJ., Rigel DS. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *NCBI*, 1985.
- [8] Crotty KA. McCarthy WH. Manzi SW., Ingvar C. Frequency and morphologic characteristics of invasive melanomas lacking specific microscopic features. *NCBI*, 1996.
- [9] Carli P. et al. Argenziano G., Fabbocini G. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new seven-point checklist based on pattern analysis.
- [10] Jacqueline Dinnes, Jonathan J Deeks, Naomi Chuchu, Lavinia Ferrante di Ruffano, Rubeta N Matin, David R Thomson, Kai Yuen Wong, Roger Benjamin Aldridge, Rachel Abbott, Monica Fawzy, et al. Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults. *Cochrane Database of Systematic Reviews*, (12), 2018.
- [11] Dermoscopia. <http://dermatologia.cat/es/informacion-para-pacientes/tecnicas-diagnosticas/dermatoscopia/>. Accessed: 2019-02-23.

- [12] Puig Susana Malveyh Josep Zaballos Pedro, Carrera Cristina. Criterios dermatoscópicos para el diagnóstico del melanoma. *Med Cutan Iber Lat Am*, 32, 2004.
- [13] Cognitive computing and artificial intelligence systems in healthcare. <https://store.frost.com/cognitive-computing-and-artificial-intelligence-systems-in-healthcare.html>.
- [14] Robert J Friedman, Dina Gutkowitz-Krusin, Michele J Farber, Melanie Warycha, Lori Schneider-Kels, Nicole Papastathis, Martin C Mihm, Paul Googe, Roy King, Victor G Prieto, et al. The diagnostic performance of expert dermoscopists vs a computer-vision system on small-diameter melanomas. *Archives of dermatology*, 144(4):476–482, 2008.
- [15] Guido Rossum. Python reference manual. Technical report, Amsterdam, The Netherlands, The Netherlands, 1995.
- [16] Lisa Torrey and Jude Shavlik. Transfer learning.
- [17] Programa Argentino de Consensos de Enfermedades Oncológicas Programa Nacional de Consensos Inter-Sociedades. Consenso Nacional Inter-Sociedades sobre melanoma cutáneo, 2011.
- [18] AJ Miller and MC Jr. Mihm. Melanoma. *NCBI*, 2006.
- [19] Sharfman William H. *Melanoma*. demos Medical, 2012.
- [20] Buliard JL et al. Van der Leest Rj, de Vries E. The Eruomelanoma skin cancer prevention campaign in Europe: characteristics and results of 2009 and 2010. *NCBI*, 2011.
- [21] Schuz J. et al. Erdamann F., Lortet-Tieulent J. International trends in the incidence of malignant melanoma 1953-2008. *NCBI*, 2012.
- [22] Dermoscopy. <https://emedicine.medscape.com/article/1130783-overview>. Accessed: 2019-03-21.
- [23] Harald Kittler, Cliff Rosendahl, Alan Cameron, and Philip Tschandl. *Dermatoscopía. Un método algorítmico basado en el análisis de patrones*. facultas, 2nd edition.
- [24] I Zalaudek, Giuseppe Argenziano, HP Soyer, R Corona, F Sera, A Blum, RP Braun, H Cabo, G Ferrara, AW Kopf, et al. Three-point checklist of dermoscopy: an open internet study. *British journal of dermatology*, 154(3):431–437, 2006.
- [25] Dermoscopedia. https://dermoscopedia.org/Blue_w_hite_v_eil. Accessed : 2019 – 04 – 25.

- [26] Dr. M. Emre Celebi, Iyatomi Dr. Hitoshi, William V. Stoecker, Dr. Randy H. Moss, Harold S. Rabinovitz, Giuseppe Argenziano, and Peter H. Soyer. Automatic Detection of Blue-White Veil and Related Structures in Dermoscopy Images.
- [27] Ali Madooei, Mark S. Drew, Maryam Sadegui, and M. Stella Arkins. Automatic detection of blue-white veil by discrete colour matching in dermoscopy images.
- [28] José Luis García Arroyo, Zapirain Begoña García, and Zorrilla Amaia Méndez. Blue-white veil and dark-red patch of pigment pattern recognition in dermoscopic images using machine-learning techniques. *IEEE*, 2011.
- [29] José Luis García Arroyo and Zapirain García. Comparison of Image Processing Techniques for Reticular Pattern Recognition in Melanoma Detection. 2015.
- [30] José Luis García Arroyo and García Zapirain Begoña. Recognition of pigment network pattern in dermoscopy images based on fuzzy classification of pixels. *Elsevier*, 2018.
- [31] Eltayef Khalid, Li Yongmin, and Liu Xiaohui. Detection of Pigment Netwirks in Dermoscopy Images.
- [32] W. V. Stoecker, W. W. Li, and R. H. Moss. Automatic detection of asymmetry in skin tumors. *NCBI*, 1992.
- [33] Stefania Seidenari, Giovanni Pellacani, and Constantino Grana. Asymmetry in Dermoscopic Melanocytic Lesion Images: a Computer Description Based on Colour Distribution. 2006.
- [34] A. Blum, H. Luedtke, U. Ellwanger, R. Schwabe, G. Rassner, and C. Garbe. Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. *NCBI*, 2004.
- [35] Stanganelli Ignazio et al. Computer-aided Diagnosis of Melanocytic Lesions. *Anti-Cancer Research*, 2005.
- [36] S. Tomatis, M. Carrara, A. Bono, C. Bartoli, M. Lualdi, G. Tragni, A. Colombo, and R. Marchesini. Automated melanoma detection with a novel multispectral imaging system: results of a prospective study. *NCBI*, 2005.
- [37] Iyatomi Hitoshi, Oka Hiroshi, M. Emre Celebi, Hashimoto Masahiro, Hagiwara Masafumi, and Masaru Tanaka Koichi Ogawa. An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *Elsevier*, 2008.

- [38] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Ann Heng Phn. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE*, 2016.
- [39] KC Nischal, Uday Khopkar, et al. Dermoscope. *Indian Journal of Dermatology, Venereology, and Leprology*, 71(4):300, 2005.
- [40] International skin image collaboration. <https://www.isic-archive.com/>. Accessed: 2019-03-04.
- [41] Gallagher R. Coldman A. McLean D. Lee T, Ng V. Dullrazor: a software approach to hair removal from images. *NCBI*, 1997 Nov. 27.
- [42] Henk J.A.M. Heijmans. *Mathematical Morphology: Basic Principles*. 1995.
- [43] L. Wiener, F. Jin, P. Fieguth, and E. Jernigan. Adaptive Wiener filtering of noisy images and image sequences. *IEEE*, 2003.
- [44] Canny John. A computational Approach to Edge Detection. *IEEE*, 1986.
- [45] Youguo Li and Haiyan Wu. A clustering Method Based on K-Means Algorithm. *Physics Procedia*, 2012.
- [46] Bhanot Karan. Color identification in images.
- [47] Maryam Sadeghi, Majid Razmara, K. Lee Tim, and M. Stella Atkins. A novel method for detection of pigment network in dermoscopic images using graphs. 35:137–143, 2011.
- [48] Andrea Polesel, Giovanni Ramponi, and V. John. Mathews. Image Enhancement via Adaptative Unsharp Masking. *IEEE*, 2000.
- [49] Skimage. <https://scikit-image.org/>. Accessed: 2018-02-10.
- [50] Samta Gupta and Mazumdar Susmita Ghosh. Sobel Edge Detection Algorithm. *International Journal of Computer Science and Management Research*, 2, 2013.
- [51] Y. LeCun, Y. Bengio, and G. Hinton. Deep Learning. *NCBI*, 2015.
- [52] An intuitive explanation of convolutional neural networks. <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>. Accessed: 2019-06-13.
- [53] Brinker Titus Josef et al. Skin cancer classification using convolutional neural networks: Systematic review. *NCBI*, 2018.
- [54] Transfer learning from pre-trained models. <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>. Accessed: 2019-05-10.

- [55] Serdar Yegulalp. What is tensorflow? the machine learning library explained. Accessed: 2019-07-15.
- [56] François Chollet et al. Keras. <https://keras.io>, 2015.
- [57] Sik-Ho Tsang. Review: Xception — With Depthwise Separable Convolution, Better Than Inception-v3 (Image Classification), 2018.
- [58] Pablo Ruiz. ResNets for CIFAR-10, 2018.
- [59] Connor Shorten. Introduction to resnets, 2019.
- [60] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. pages 770–778, 2016.
- [61] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- [62] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [63] Shie Mannor, Dori Peleg, and Reuven Rubinfeld. The cross entropy method for classification. pages 561–568, 2005.
- [64] GM Foody, MB McCulloch, and WB Yates. The effect of training set size and composition on artificial neural network classification. *International Journal of Remote Sensing*, 16(9):1707–1723, 1995.