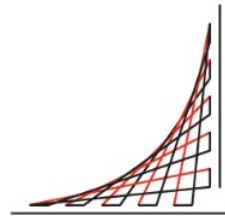




Universidad del  
**Rosario**



ESCUELA  
COLOMBIANA  
DE INGENIERÍA  
JULIO GARAVITO

# Detección de melanoma cutáneo a través de la regla ABCD usando algoritmos de aprendizaje profundo

**Paula Caterine Moreno Luna**

**Tutor**

**Ph.D. Oscar Julián Perdomo Charry**

**Ph.D. Sebastián Jaramillo Isaza**

UNIVERSIDAD DEL ROSARIO  
UNIVERSIDAD ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO  
PROGRAMA DE INGENIERÍA BIOMÉDICA  
BOGOTÁ, COLOMBIA

2023

# Detección de melanoma cutáneo a través de la regla ABCD usando algoritmos de aprendizaje profundo

Paula Caterine Moreno Luna

Trabajo final de maestría presentado como requisito para optar al título de:  
**Magister en Ingeniería Biomédica**

Tutor:

Ph.D. Oscar Julián Perdomo Charry

Ph.D. Sebastián Jaramillo Isaza

UNIVERSIDAD DEL ROSARIO  
UNIVERSIDAD ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO  
PROGRAMA DE INGENIERÍA BIOMÉDICA  
BOGOTÁ, COLOMBIA  
2023

*”Lo que con mucho trabajo se obtiene, más se ama”*

Aristóteles

# Agradecimientos

En primer lugar, me gustaría agradecer a mis papas y mis hermanos por su apoyo y guía durante este proceso. A mis papas quiero darles las gracias por enseñarme el valor del aprendizaje y la importancia de querer salir adelante, a mi hermana por su amistad, compañía constante y apoyo incondicional. También me gustaría agradecer a mis profesores por su guía y enseñanza constante a lo largo de esta maestría, a mi tutor, el profesor Oscar Perdomo, por su enseñanza, compañía, paciencia, dedicación, compromiso y orientación en este camino. Finalmente, agradecer a todas las personas involucradas en todo este proceso, su apoyo, compañía y colaboración permitieron que este proyecto culmine con éxito.

## Resumen

El melanoma es un tipo de cáncer de piel que afecta los melanocitos, estas son las células encargadas de la producción de melanina, esta neoplasia es la de mayor mortalidad de todos los cánceres de piel. Durante los últimos años se ha venido presentando un aumento de nuevos casos, por ejemplo, en Colombia las cifras dadas por la Cuenta de Alto Costo lo ubicaron en 2021 como el octavo tumor más frecuente según el número de diagnósticos registrados, y la mortalidad aumentó un 30% en comparación con el año anterior. Hoy en día existen una gran cantidad de métodos y técnicas para identificarlo en sus etapas tempranas, siendo una de ellas la regla ABCD, la cual a través de las características físicas del lunar puede determinar la sospecha de que este tenga células cancerosas, de manera específica: Si el lunar es asimétrico (A), tiene bordes irregulares (B), más de uno o dos colores (C) y un diámetro mayor a 6 mm (D) tiene una alta probabilidad de que sea melanoma, esta regla ha ganado bastante aceptación clínica para la identificación de esta enfermedad. Con base en esto, el objetivo de este trabajo de maestría fue la adaptación de modelos basados en aprendizaje profundo para la estimación automática de características que puedan clasificar lunares como benignos o malignos, con validación en imágenes obtenidas de la base de datos del *International Skin Imaging Collaboration (ISIC) Challenge Dataset*. Para ello se usaron técnicas clásicas de procesamiento de imágenes para calcular las características ABCD de toda la base de datos y luego se realizó el entrenamiento para clasificación usando: las características, las imágenes y ambos conjuntos de datos, con esto se obtuvo que aquellos modelos donde se usaron tanto las características como las imágenes tuvieron un mejor desempeño, alcanzando mejores métricas de desempeño tanto en entrenamiento como en prueba.

### **Palabras clave.**

Aprendizaje de máquina, Aprendizaje profundo, Clasificación, Melanoma, Regla ABCD.

# Abstract

Melanoma is a type of skin cancer that affects melanocytes, which are the cells responsible for the production of melanin. This neoplasm has the highest mortality rate of all skin cancers. During recent years there has been an increase in new cases, for example, in Colombia the figures given by the High Cost Account placed it as the eighth most frequent tumor in 2021 according to the number of registered diagnoses, and mortality increased by 30% compared to the previous year. Nowadays there are a large number of methods and techniques to identify it in its early stages, one of them being the ABCD rule, which through the physical characteristics of the mole can determine the suspicion that it has cancerous cells, specifically. : If the mole is asymmetrical (A), has irregular edges (B), more than one or two colors (C), and has a diameter greater than 6 mm (D) it has a high probability of being melanoma, this rule has gained a lot clinical acceptance for the identification of this disease. Based on this, the objective of this master's work was the adaptation of models based on deep learning for the automatic estimation of characteristics that can classify moles as benign or malignant, with validation in images obtained from the *database International Skin Imaging Collaboration (ISIC) Challenge Dataset*. For this, classic image processing techniques were used to calculate the ABCD characteristics of the entire database and then the training task was systematically assessed using: only the characteristics, only the images, and the combination of both sources of information (features and images), with this final approach, was obtained that both features and images performed better performance metrics in training and testing sets.

**Key words.**

ABCD rule, Classification, Deep learning, Machine learning, Melanoma.

# Tabla de Contenidos

|  |           |
|--|-----------|
| <b>Agradecimientos</b>   | <b>iv</b> |
| <b>Resumen</b>   | <b>v</b>  |
| <b>Abstract</b>  | <b>vi</b> |
| <b>Índice de Figuras</b>   | <b>ix</b> |
| <b>Índice de Tablas</b>  | <b>xi</b> |
| <b>1 INTRODUCCIÓN</b>  | <b>1</b>  |
| 1.1 Organización del documento . . . . .   | 2         |
| <b>2 PROBLEMA DE INVESTIGACIÓN</b>   | <b>3</b>  |
| 2.1 Antecedentes . . . . .   | 3         |
| 2.2 Descripción del problema . . . . .   | 5         |
| 2.3 Justificación . . . . .  | 7         |
| <b>3 OBJETIVOS</b>   | <b>9</b>  |
| 3.1 Objetivo General . . . . .   | 9         |
| 3.2 Objetivos Específicos . . . . .  | 9         |
| <b>4 MARCO TEÓRICO</b>   | <b>10</b> |
| 4.1 Melanoma . . . . .   | 10        |
| 4.2 Base de datos ISIC Challenge . . . . .   | 12        |
| 4.3 Inteligencia Artificial . . . . .  | 13        |
| 4.3.1 Aprendizaje de máquina . . . . .   | 14        |
| 4.3.2 Aprendizaje profundo . . . . .   | 17        |
| <b>5 METODOLOGÍA</b>   | <b>20</b> |
| 5.1 Actividad preliminar: Identificación de la base de datos, selección de la información y almacenamiento . . . . . | 20        |
| 5.2 Fase I: Obtención de las características de los lunares con base en la regla ABCD. . . . .                       | 21        |
| 5.2.1 Asimetría . . . . .  | 22        |

---

|          |   |           |
|----------|---|-----------|
| 5.2.2    | Bordes irregulares . . . . .  | 22        |
| 5.2.3    | Colores . . . . .   | 23        |
| 5.2.4    | Diámetro . . . . .  | 24        |
| 5.3      | Fase II: Adaptación de modelos de aprendizaje profundo para la extracción de características y clasificación de melanoma de las imágenes dermatoscópicas de lunares . . . . . | 24        |
| 5.3.1    | Adaptación de los modelos para clasificar melanoma usando las características ABCD extraídas en la Fase I . . . . .   | 25        |
| 5.3.2    | Adaptación de los modelos para clasificar melanoma usando las imágenes de la base de datos . . . . .  | 27        |
| 5.3.3    | Adaptación de los modelos para clasificar melanoma usando las imágenes de la base de datos y el archivo de características ABCD . . . . .                                     | 27        |
| 5.4      | Fase III: Validación de la habilidad del método para clasificar melanoma usando las imágenes de la base de datos ISIC. . . . .  | 28        |
| <b>6</b> | <b>RESULTADOS</b>   | <b>30</b> |
| 6.1      | Resultados de la obtención de las características ABCD . . . . .  | 30        |
| 6.2      | Resultados del entrenamiento usando las características ABCD . . . . .  | 35        |
| 6.3      | Resultados del entrenamiento usando las imágenes de la base de datos ISIC . . . . .   | 42        |
| 6.4      | Resultados del entrenamiento usando las imágenes de la base de datos ISIC junto con sus características ABCD . . . . .  | 46        |
| <b>7</b> | <b>DISCUSIÓN</b>  | <b>55</b> |
| <b>8</b> | <b>RECOMENDACIONES Y TRABAJOS FUTUROS</b>   | <b>59</b> |
| <b>9</b> | <b>CONCLUSIONES</b>   | <b>60</b> |
|          | <b>Referencias</b>  | <b>62</b> |

# Índice de Figuras

|     |  |    |
|-----|--|----|
| 2-1 | Tendencia del melanoma en Colombia entre 2015 - 2021. Disponible en [17] .                               | 7  |
| 4-1 | Capas de la piel humana. Disponible en [46] . . . . .  | 10 |
| 4-2 | Ejemplo de un lunar con melanoma tomado de la base de datos ISIC. Disponible en [27] . . . . .           | 11 |
| 4-3 | Regla ABCD para detectar melanoma. Disponible en [37] . . . . .  | 13 |
| 4-4 | Gráfica representando la función empleada para regresión logística. Disponible en [51] . . . . .         | 15 |
| 4-5 | Representación gráfica del modelo KNN. Disponible en[3] . . . . .  | 16 |
| 4-6 | Gráfica de la función ReLU. Disponible en[47] . . . . .  | 18 |
| 4-7 | Representación gráfica de la arquitectura InceptionResNet. Disponible en[31]                             | 19 |
| 4-8 | Representación gráfica de la arquitectura EfficientNet. Disponible en[53] . .                            | 19 |
| 5-1 | Diagrama general de metodología . . . . .  | 20 |
| 5-2 | Máscara del lunar con su correspondiente bounding box . . . . .  | 22 |
| 5-3 | Las cuatro mitades halladas para encontrar el valor de asimetría . . . . .                               | 23 |
| 5-4 | Imagen obtenida de la multiplicación de la imagen a color y su respectiva máscara . . . . .              | 24 |
| 5-5 | Comparación lunar benigno y uno con melanoma. Imágenes disponibles en [26].                              | 25 |
| 6-1 | Diagrama de cajas del conjunto de entrenamiento para la característica de asimetría horizontal . . . . . | 31 |
| 6-2 | Diagrama de cajas del conjunto de prueba para la característica de asimetría horizontal . . . . .        | 31 |
| 6-3 | Diagrama de cajas del conjunto de entrenamiento para la característica de asimetría vertical . . . . .   | 32 |
| 6-4 | Diagrama de cajas del conjunto de prueba para la característica de asimetría vertical . . . . .          | 32 |
| 6-5 | Diagrama de cajas del conjunto de entrenamiento para la característica de bordes . . . . .               | 33 |
| 6-6 | Diagrama de cajas del conjunto de entrenamiento para la característica de cantidad de colores . . . . .  | 34 |
| 6-7 | Diagrama de cajas del conjunto de entrenamiento para la característica de diámetro . . . . .             | 35 |

|             |   |    |
|-------------|---|----|
| <b>6-8</b>  | Gráficas de dispersión del conjunto de entrenamiento para cuatro de las siete características obtenidas . . . . .                                     | 36 |
| <b>6-9</b>  | Gráfica de densidad para dos pares de características (Asimetría horizontal con colores en el canal rojo y asimetría vertical con diámetro) . . . . . | 38 |
| <b>6-10</b> | Matriz de confusión del modelo de regresión logística de las imágenes de prueba   | 39 |
| <b>6-11</b> | Matriz de confusión del modelo KNN de las imágenes de prueba . . . . .  | 40 |
| <b>6-12</b> | Matriz de confusión del modelo Random Forest de las imágenes de prueba .  | 41 |
| <b>6-13</b> | Matriz de confusión del modelo de máquinas de soporte vectorial de las imágenes de prueba . . . . .   | 42 |
| <b>6-14</b> | Matrices de confusión de las tres configuraciones del modelo perceptrón multicapa para el conjunto de entrenamiento . . . . .                         | 43 |
| <b>6-15</b> | Tabla de los mejores modelos mostrada luego del uso de la librería <i>lazy predict</i>  | 44 |
| <b>6-16</b> | Matriz de confusión del modelo Nearest Centroid . . . . .   | 46 |
| <b>6-17</b> | Matrices de confusión de los otros tres mejores modelos de <i>Lazy Predict</i> . . .  | 47 |
| <b>6-18</b> | Matriz de confusión de la etapa de congelamiento de capas del conjunto de entrenamiento 2017 . . . . .  | 48 |
| <b>6-19</b> | Matriz de confusión de la segunda etapa del conjunto de entrenamiento de 2016   | 49 |
| <b>6-20</b> | Matriz de confusión de la segunda etapa del conjunto prueba de 2017 . . . . .   | 50 |
| <b>6-21</b> | Matriz de confusión de la tercera etapa del conjunto de entrenamiento de 2017   | 51 |
| <b>6-22</b> | Matriz de confusión de la tercera etapa del conjunto prueba de 2016 . . . . .   | 51 |
| <b>6-23</b> | Matriz de confusión de la arquitectura InceptionResNet con una capa adicional del conjunto de entrenamiento . . . . .                                 | 52 |
| <b>6-24</b> | Matriz de confusión de la arquitectura InceptionResNet con dos capas adicionales del conjunto de prueba . . . . .                                     | 52 |
| <b>6-25</b> | Matriz de confusión de la arquitectura InceptionResNet con tres capas adicionales del conjunto de prueba . . . . .                                    | 53 |
| <b>6-26</b> | Matriz de confusión de la arquitectura EfficientNet con una capa adicional del conjunto de prueba . . . . .   | 53 |
| <b>6-27</b> | Matriz de confusión de la arquitectura EfficientNet con dos capas adicionales del conjunto de entrenamiento . . . . .                                 | 54 |
| <b>6-28</b> | Matriz de confusión de la arquitectura EfficientNet con tres capas adicionales del conjunto de prueba . . . . .                                       | 54 |

# Índice de Tablas

|            |  |    |
|------------|--|----|
| <b>2-1</b> | Resumen de los resultados de trabajos relacionados (Exac: Exactitud – Prec: Precision - Sens: Sensibilidad – Espe: Especificidad ) . . . . . | 6  |
| <b>5-1</b> | Parámetros empleados explorar usando GridSearch . . . . .  | 26 |
| <b>5-2</b> | Mejores combinaciones como resultado de emplear GridSearch . . . . .   | 27 |
| <b>5-3</b> | Interpretación del coeficiente Kappa. Disponible en [44] . . . . .   | 29 |
| <b>6-1</b> | Métricas obtenidas en los 5 modelos clásicos en fase de entrenamiento . . . . .  | 37 |
| <b>6-2</b> | Métricas obtenidas en los 5 modelos clásicos en fase de prueba . . . . .   | 37 |
| <b>6-3</b> | Métricas de desempeño de los 4 mejores modelos de <i>Lazy Predict</i> . . . . .  | 41 |
| <b>6-4</b> | Métricas de desempeño para el conjunto de entrenamiento de la base de datos del año 2016 . . . . .   | 44 |
| <b>6-5</b> | Métricas de desempeño para el conjunto de prueba de la base de datos del año 2016 . . . . .  | 45 |
| <b>6-6</b> | Métricas de desempeño para el conjunto de entrenamiento de la base de datos del año 2017 . . . . .   | 45 |
| <b>6-7</b> | Métricas de desempeño para el conjunto de prueba de la base de datos del año 2017 . . . . .  | 45 |
| <b>6-8</b> | Métricas de desempeño para el conjunto de entrenamiento de la base de datos . . . . .  | 48 |
| <b>6-9</b> | Métricas de desempeño para el conjunto de prueba de la base de datos . . . . .   | 49 |
| <b>7-1</b> | Tabla comparativa de los mejores modelos en cada fase para entrenamiento . . . . .   | 58 |
| <b>7-2</b> | Tabla comparativa de los mejores modelos en cada fase para prueba . . . . .  | 58 |



# 1 INTRODUCCIÓN

El melanoma cutáneo es un tipo de cáncer de piel que se desarrolla en los melanocitos [48], las células encargadas de la piel que producen y almacenan la melanina [15]. Esta es la neoplasia de piel con mayor mortalidad, con una tasa de 3.1 fallecimientos por cada 100000 personas en los Estados Unidos [19], generalmente se encuentra en lunares de pacientes con piel clara [50] siendo este uno de los factores de riesgo del melanoma [41], fenotipos como los ojos y/o el cabello claro aumentan el riesgo de padecer de esta enfermedad, un metaanálisis realizado en varios estudios encontró que los pacientes con cabello rojo tienen 3.64 más probabilidades de aquellos que tienen el cabello oscuro [41].

Sin embargo, el mayor factor de riesgo que existe, es la exposición a los rayos UV, clasificada como un carcinógeno de tipo 1 [25], siendo los rayos UVB responsables de no solo los bronceados duraderos [45] sino también de causar daño en el ADN a través de foto productos capaces de causar esto [32], mientras que los rayos UVA causan el bronceado inmediato [45] y dañan de manera indirecta el ADN al ocasionar la producción de especies reactivas de oxígeno [32].

Debido a la capacidad de este cáncer de hacer metástasis [41], realizar el diagnóstico a tiempo es de vital importancia, puesto que puede reducir la mortalidad [18]. Para obtenerlo en primer lugar, se requiere de la identificación de un área, generalmente un lunar, que tenga con un aspecto sospechoso, esto lo puede realizar el mismo paciente o un médico bien sea que este tenga experiencia o no en la identificación de estas lesiones, conocido como la examinación visual [41], luego de ello se realiza la respectiva biopsia para confirmar si el tejido resulta ser benigno o maligno [18]. Sin embargo, en Colombia el melanoma no suele detectarse a tiempo: aproximadamente 4.5 diagnósticos por cada 100000 pacientes al año [52], y en caso de tener un diagnóstico los tiempos de espera para el inicio del tratamiento pueden llegar a ser de hasta 100 días [17].

Por este motivo, se propone la realización de este proyecto: desarrollar una herramienta que puede llegar a ser un apoyo diagnóstico, haciendo uso de diferentes técnicas de aprendizaje profundo, puede ayudar a aquellos pacientes quienes tienen historial de antecedentes familiares con melanoma, o a los médicos que no tienen la experiencia en reconocer estas lesiones. Esto puede beneficiar a todas las partes involucradas, pues al aumentar el número de diagnósticos en Colombia al año, el melanoma puede llegar a estar en las etapas tempranas

de la enfermedad, por lo cual los gastos involucrados en el tratamiento serían menores. Así mismo, una herramienta de apoyo diagnóstico puede ayudar a disminuir los tiempos de espera tanto para el diagnóstico como para el inicio del tratamiento; todo esto ayudaría a disminuir las tasas elevadas de mortalidad que presenta el melanoma.

## 1.1 Organización del documento

En este documento se presentan 9 capítulos distribuidos y descritos a continuación:

- Capítulo 2: Se describe el problema de investigación incluyendo los antecedentes de estudios relacionados a este proyecto, la descripción del problema de investigación y la justificación de este mismo.
- Capítulo 3: Se exponen tanto el objetivo general, como todos los objetivos específicos de este proyecto de investigación.
- Capítulo 4: Se da una explicación acerca de los temas relevantes para esta investigación como lo es el melanoma, la regla ABCD, los modelos de *Machine Learning* clásicos empleados, y las arquitecturas exploradas para la clasificación de las imágenes.
- Capítulo 5: Se explica de manera detalla el paso a paso realizado en cada una de las fases de la metodología, qué se hizo, cómo se hizo.
- Capítulo 6: Se exponen los resultados obtenidos a lo largo del desarrollo de esta investigación, como lo es la obtención de las características ABCD para cada una de las imágenes de la base de datos, las métricas de desempeño calculadas para cada modelo de *Machine Learning* adaptado a la tarea de clasificación: con las características ABCD, con las características extraídas automáticamente por los modelos y con la combinación de ambos conjuntos.
- Capítulo 7: Se realiza una discusión de resultados, con base en los resultados expuestos en el capítulo anterior, y se compara con los resultados obtenidos por estudios similares de los últimos 10 años.
- Capítulo 8: Se da una explicación de las recomendaciones y trabajos futuros que surgen con base en lo desarrollado y los resultados obtenidos con este proyecto de investigación.
- Capítulo 9: Conclusiones del proyecto, teniendo en cuenta lo que se propone en los objetivos y los resultados obtenidos.

# 2 PROBLEMA DE INVESTIGACIÓN

En este capítulo se brindará el contexto alrededor del problema de investigación de este proyecto: la dificultad que presentan los médicos de atención primaria a la hora de diagnosticar melanoma. Para ello, en primer lugar, se darán los antecedentes en los cuales aborden este problema por medio del uso de las diferentes técnicas de inteligencia artificial, seguido de la descripción del problema para finalmente brindar la justificación alrededor de este. La información que se muestra a continuación es el producto de una revisión literaria de artículos académicos obtenidos de fuentes de información como: *Science Direct*, *Scopus*, *Google Scholar*, *Pubmed*, *SciELO*, entre otros.

## 2.1 Antecedentes

Según la revisión de literatura realizada en 2020 por Ali et al. [2], quienes primero incurrieron en la detección de melanoma usando aprendizaje profundo fueron Codella et al. [10] quienes combinaron técnicas de aprendizaje profundo, con máquinas de soporte vectorial (*Support Vector Machine*, SVM) y codificación dispersa para detectar melanoma de imágenes dermatoscópicas de la base de datos *International Skin Imaging Collaboration* (ISIC). Para ello usaron redes neuronales convolucionales y aprendizaje no supervisado usando codificación dispersa para la extracción de características, posterior a esto se entrenan las SVM para luego combinar todos los modelos usando *late fusion* [10]; de esto se obtiene para las redes neuronales una exactitud de hasta el 91.9%, con una sensibilidad del 90.3% y una especificidad de 92.1%; además de los resultados reportadas, el uso de *late fusión* fue el que mejor resultados presento, una exactitud de 93.1% [10].

A partir de este momento, la cantidad de estudios que usan aprendizaje profundo para detectar melanoma ha aumentado significativamente, según la revisión mencionada [2] entre el periodo de 2015 y 2017 las publicaciones que relacionan estos temas aumento hasta un 100%, sin embargo, desde 2018 el tener control de este dato ha presentado mayores dificultades debido a la gran cantidad de artículos publicados sobre esto. Por ejemplo, Moura et al [34] extrajeron características de las imágenes segmentadas usando la regla ABCD junto con las extraídas de redes neuronales convolucionales pre entrenadas, para luego seleccionarlas con base en un umbral y con ella realizar la clasificación de las imágenes usando un perceptrón multicapa, de esto se obtuvo que para el mejor resultado obtenido, que consiste en el uso de varios descriptores como la regla ABCD, las arquitecturas Vgg y CaffeNet junto con el

perceptrón multi capa como clasificador, la exactitud alcanzo un 94,9% y una precisión de 94,8% junto con un índice Kappa de 89.2% [34], esto quiere decir que el acuerdo entre el modelo propuesto y lo establecido por el experto es excelente.

En el 2018 Hardie et al.[24] publicaron el estudio realizado durante el reto de ISIC donde emplearon el conjunto de imágenes de este mismo año; para la tarea de clasificación usaron maquinas de soporte vectorial con 200 características extraídas a mano de siete lesiones de piel estas se obtuvieron con base en estudios previos donde se detectaban nódulos pulmonares. Para las imágenes de la clase melanoma se obtuvo una exactitud de 87.08%, con sensibilidad de 70.35% y especificidad de 89.17% [24]. De manera adicional, existen metodologías como la propuesta por Poorna et al. [40] en la cual desean comparar diferentes métodos de clasificación de melanoma como: el puntaje dermatoscópico total, que usa la regla ABCD; los clasificadores clásicos como el *K-Nearest Neighbors* (KNN) y el SVM; y las redes neuronales convolucionales (*Convolutional neural network*, CNN), al realizar la comparación de resultados de estos tres métodos diferentes, obtuvieron que el menor valor de exactitud fue para el puntaje dermatoscópico total correspondiente a 57.3%, mientras que la CNN probada, que fue con la arquitectura AlexNet, tuvo una exactitud de 90% y una precisión de 83.78% [40].

En 2020, Kassem et al [28] usando aprendizaje profundo, proponen usar *transfer learning* y un modelo de GoogleNet pre entrenado para clasificar 8 diferentes lesiones de la piel, dentro de las que se encuentra el melanoma, en este estudio realizaron dos experimentos: el primero usaron las imágenes de la base de datos, ISIC, obteniendo una exactitud de 93.31%, pero una sensibilidad de 53.3%, esto debido a que cada clase, respecto al número de imágenes se encontraba desbalanceado [28]. Razón por la cual se propuso un aumento del número de imágenes para balancear la proporción entre clases y eliminar el sesgo que presentaba el modelo, con esto la exactitud alcanzo un 94.2% y la sensibilidad aumento a 74.5% [28]. El año siguiente Yilmaz et al. [54] publicaron el artículo titulado *Benchmarking of Lightweight Deep Learning Architectures for Skin Cancer Classification using ISIC 2017 Dataset* donde muestran los resultados obtenidos al entrenar 9 modelos diferentes, tres CNN junto con tres tamaños de lote distintos, con esto se obtuvo que la red neuronal convolucional NASNet-Mobile entrenada con un tamaño de lote de 16 mostro los mejores resultados al tener una exactitud de 82% y una precisión de 81.77% [54].

Hablando de investigaciones más recientes, en 2022, Salma et al [43] proponen un sistema de diagnóstico asistido por computador en el cual se realiza un preprocesamiento para eliminar el vello y demás artefactos, seguido de una segmentación de las imágenes y con esto emplear la regla ABCD de manera automática para realizar la clasificación de los lunares, usando diferentes CNN pre entrenadas. Para la extracción de las características ABCD lo hacen de la siguiente manera [43]:

- **Asimetría:** Dos ejes perpendiculares que crucen el centro de la imagen y se determina la asimetría de la imagen por cada eje para el color, brillo y forma.
- **Bordes:** La imagen se divide en 8 partes, y se evalúa cada parte para determinar la irregularidad del borde si este presenta un corte brusco o gradual.
- **Color:** Se determina que si el área de interés de la imagen, el lunar, tiene más de un color cuando el número de píxeles de ese color sobrepasa por 5% el total de píxeles del área evaluada.
- **Estructuras dermatoscópicas:** Evalúan la existencia de estructuras como: Red, Áreas sin estructura, Estrías ramificadas, Puntos y Glóbulos dentro de la imagen.

Con esto se realiza la clasificación de las imágenes usando varias CNN pre entrenadas con los pesos de *ImageNet*, y en ocasiones, combinándolas con una SVM, a partir de esto se obtuvo que los mejores resultados fueron con la arquitectura ResNet50 con SVM que tuvo una exactitud de 99.87% y una precisión de 98.77% [43].

Por otro lado, en este mismo año Benyahia et al. [5] publicaron los resultados del estudio en donde evalúan la eficiencia de 17 redes neuronales convoluciones pre entrenadas para la extracción de características, así como de 24 clasificadores basados en aprendizaje de máquina para la clasificación de diferentes lesiones de la piel, incluida el melanoma. A partir de esto se obtuvo que la CNN DenseNet201 usada en conjunto con el clasificador fino de KNN mostro las mejores métricas de desempeño con exactitud de 92.34%, precisión de 85.22% y especificidad de 96.38% [5]. Toda la información expuesta anteriormente se encuentra resumida en la tabla **2-1**

## 2.2 Descripción del problema

Con base en las investigaciones revisadas, se observa una tendencia de ellas a realizar clasificación de melanoma y otras lesiones de la piel usando directamente la extracción de características que hacen las CNNs y en ocasiones usan diferentes técnicas para mejorar el rendimiento como *fine tuning*, o *data augmentation*. Sin embargo, no existen muchos estudios que usen las características de la regla ABCD bien sea para emplearlas junto con métodos de aprendizaje profundo y realizar la clasificación o para lograr la automatización de esta regla. Además, si bien la base de datos empleada en este proyecto, ISIC, ha sido muy empleada desde que la publicación del primer reto, esta suele emplearse más para tareas de segmentación de las imágenes y para detección de estructuras dermatoscópicas como redes, glóbulos, puntos, estrías ramificadas, entre otras. No obstante, es reducido el número de investigaciones que emplean la regla ABCD, junto con las imágenes para hacer un modelo

| Año publicación | Autores         | Lugar investigación                      | Métodos   | Métricas obtenidas   |
|-----------------|-----------------|--|---|--|
| 2015            | Codella et al.  | NY, USA - Merbourne, Australia           | Técnicas aprendizaje profundo con máquinas de soporte vectorial y codificación dispersa             | <b>Exac:</b> 91,9%<br><b>Sens:</b> 90,3%                               |
| 2018            | Moura et al.    | Teresina, Brasil - Picos, Brasil         | Extracción características con ABCD Y CNN preentrenadas y clasificación usando perceptrón multicapa | <b>Exac:</b> 94,9%<br><b>Prec:</b> 94,8%<br><b>Indice Kappa:</b> 89,2% |
| 2018            | Hardie et al.   | Dayton, USA                              | SVM con características extraídas a mano  | <b>Exac:</b> 87,08%<br><b>Sens:</b> 70,35%<br><b>Espec:</b> 89,17%     |
| 2020            | Poorna et al.   | Amritapuri, India                        | Comparación de puntaje dermatoscópico total, KNN, SVM y CNN   | <b>Exac:</b> 90%<br><b>Prec:</b> 83,78%                                |
| 2020            | Kassem et al    | Kafr El-Sheikh, Egipto - Zagazig, Egipto | Aprendizaje profundo con transfer learning  | <b>Exac:</b> 94,2%<br><b>Sens:</b> 74,5%                               |
| 2021            | Yilmaz et al.   | Istanbul, Turquía                        | CNN con diferentes tamaños de lote  | <b>Exac:</b> 82%<br><b>Prec:</b> 81.77%                                |
| 2022            | Salma et al     | Alexandria, Turkey                       | CAD con CNN pre entrenadas  | <b>Exac:</b> 99,87%<br><b>Prec:</b> 98,77%                             |
| 2022            | Benyahia et al. | Mascara, Argelia - Caen, Francia         | CNN para extracción de características y clasificadores basados en aprendizaje de máquina           | <b>Exac:</b> 92.34%<br><b>Prec:</b> 85.22%<br><b>Espec:</b> 96.38%.    |

**Tabla 2-1:** Resumen de los resultados de trabajos relacionados (Exac: Exactitud – Prec: Precision - Sens: Sensibilidad – Espe: Especificidad )

multimodal y hacer la clasificación.

Por otro lado, dentro de la revisión de literatura realizada, a nuestro conocimiento no se encontraron investigaciones relacionadas a este proyecto que hayan sido desarrolladas en Colombia. De aquí la importancia de fomentar estos estudios dentro del contexto colombiano para el desarrollo de tecnología que brinde el apoyo diagnóstico para esta enfermedad y poder brindarla a todos aquellos médicos que dentro de su contexto de su contexto profesional necesitan de estas herramientas bien sea por falta de experiencia con el reconocimiento de estas lesiones o por los recursos con los que cuenta su lugar de trabajo.

Con base en la información suministrada, se identifica el problema que se busca resolver con esta tesis de maestría, la cual busca emplear técnicas de aprendizaje profundo adaptadas para la extracción de características y la clasificación automática de imágenes dermatoscópicas de lunares para así brindar el punto de partida para el desarrollo de investigaciones futuras que buscan el crear una herramienta que pueda ser empleada en un contexto clínico para fomentar el diagnóstico del melanoma dentro de sus etapas tempranas.

Esta investigación es importante ya que su enfoque no se basa únicamente en la correcta clasificación automática de melanoma, sino también en el poder brindar una herramienta de base que busca el tener más diagnósticos tempranos para poder, como consecuencia, tener tratamientos que tienen mayor probabilidad de éxito dentro de las primeras etapas de este

cáncer, y así reducir las tasas de mortalidad.

## 2.3 Justificación

El melanoma cutáneo ha venido aumentando su incidencia de manera significativa en los últimos años, siendo responsable de aproximadamente el 80% de los fallecimientos asociados a cáncer de piel [1], a pesar de ser el tumor con menor frecuencia de diagnóstico [7]. Según los datos de Globocan, para el año 2020 se presentaron más de 300 mil nuevos casos a nivel mundial y casi 60 mil muertes [8]. Sin embargo, a pesar de que esta neoplasia no es frecuentemente diagnosticada en Colombia, lo que hace que los datos recopilados no sean fieles a la realidad que presenta esta enfermedad en el país [52], se estima que solo están siendo diagnosticados 4.5 de cada 100 mil pacientes al año [52], y según la Cuenta de Alto Costo a nivel nacional en el año 2020 se presentaron 6 mil nuevos casos ubicándolo como el octavo tumor más frecuente [17] y, observando los datos registrados en 2015, como se pueden ver en la figura 2-1, se observó un aumento importante en esta cifra, especialmente en el 2021, donde se registró un aumento del 30% de la mortalidad en comparación con el año anterior [17].

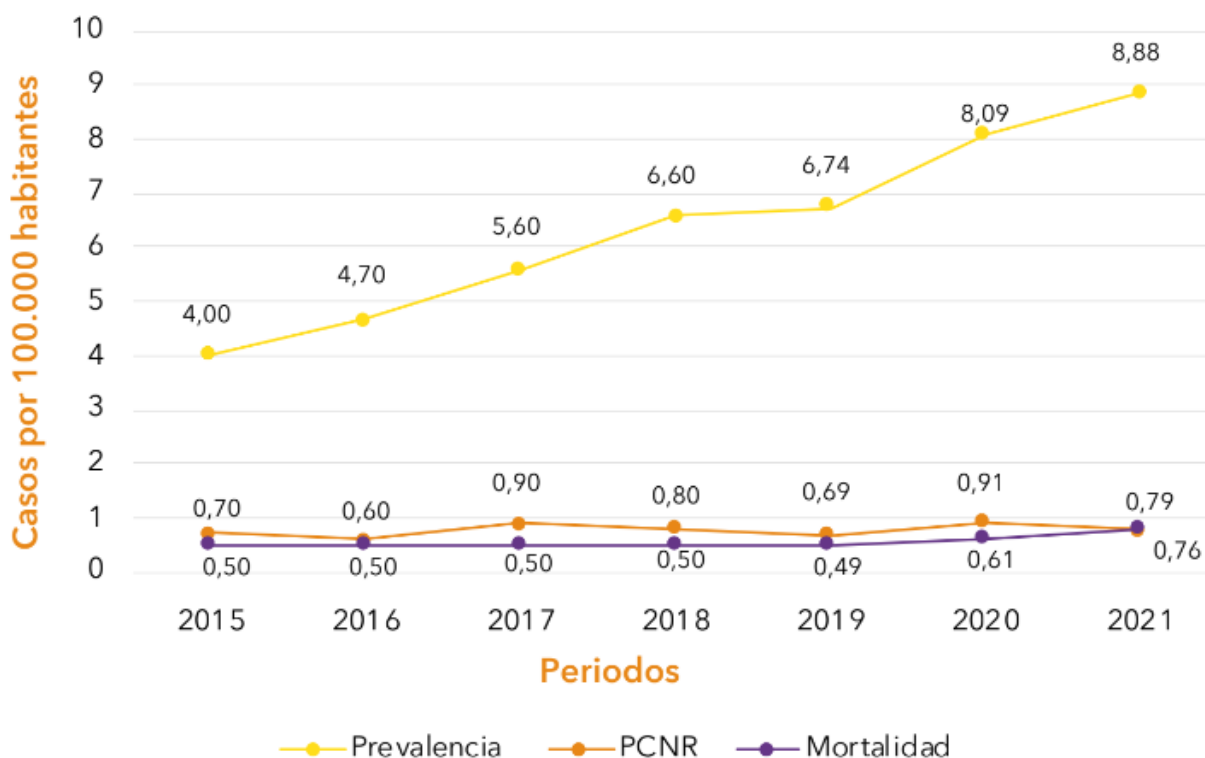


Figura 2-1: Tendencia del melanoma en Colombia entre 2015 - 2021. Disponible en [17]

Siendo la mayoría de los casos encontrados en mujeres mayores de 60 años [17] ubicadas en la región central y en la capital del país, con un 39.5% de los tumores ubicados en zonas de la piel con alta exposición a la radiación UV [9]. De manera adicional, respecto a los tiempos de espera a nivel nacional se encuentra que el tiempo entre la toma de la biopsia y el resultado de la misma fueron 11.35 días, mientras que entre el resultado esta y el inicio del tratamiento fue de 79.92 días [17], sin embargo, no se presentan datos para otros indicadores expuestos en el Consenso Basado En La Evidencia Parala Definición De Indicadores De Gestión Del Riesgo En Pacientes Con Melanoma Cutáneo En Colombia como el tiempo de espera de los pacientes para ser remitidos desde atención primaria a dermatólogo especialista [7], pero es mencionado la importancia de mejorar los tiempos de espera en atención.

De esta manera, es evidente que, a pesar de los esfuerzos por parte de las entidades competentes en cumplir con la meta de los indicadores, es importante incluir diferentes herramientas que permitan ayudar al cumplimiento de estas metas, que facilitan el camino para tener diagnósticos a tiempo y en etapas tempranas donde la supervivencia a 10 años alcanza el 85%. Investigar acerca de los métodos de clasificación automática de imágenes dermatoscópicas puede ser un camino por tomar para suplir esta necesidad que permita ayudar a mejorar la calidad de vida de aquellas personas que tienen mayor predisposición a desarrollar melanoma cutáneo.

# 3 OBJETIVOS

## 3.1 Objetivo General

Emplear técnicas de aprendizaje profundo para la extracción de características morfológicas y clasificación automática de imágenes dermatoscópicas de lunares de la base de datos *ISIC Challenge* para el apoyo diagnóstico de melanoma cutáneo.

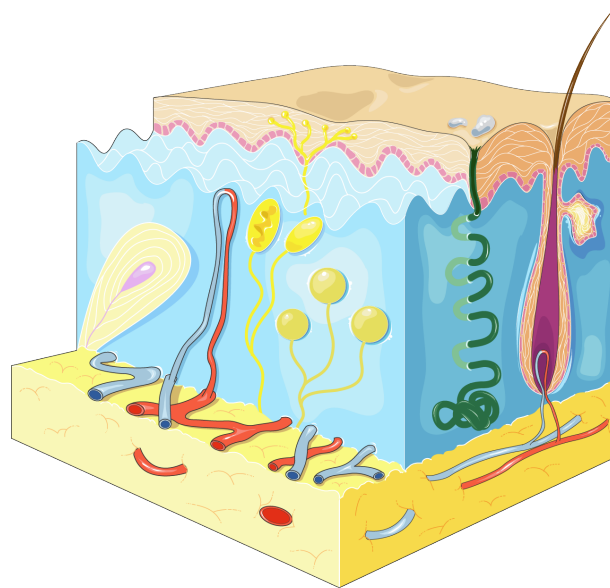
## 3.2 Objetivos Específicos

- Obtener características de asimetría, bordes irregulares, número de colores y diámetro empleadas en la regla ABCD en imágenes dermatoscópicas de lunares.
- Adaptar modelos de aprendizaje profundo para la extracción de características y clasificación de melanoma de las imágenes dermatoscópicas de lunares.
- Validar la habilidad del método propuesto para la clasificación de melanoma usando métricas como exactitud, precisión, sensibilidad y especificidad.

# 4 MARCO TEÓRICO

## 4.1 Melanoma

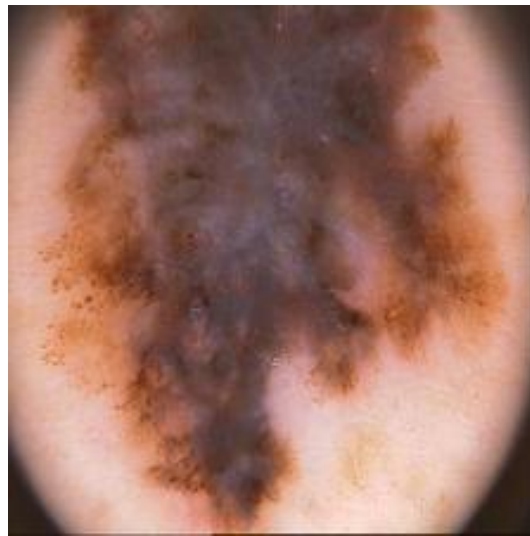
Para entender el melanoma cutáneo es importante comprender el tejido en el cual se desarrolla, la piel. Esta cuenta con varias capas, cómo se puede ver en la figura 4-1, siendo la más externa la epidermis, encargada de prohibir la entrada de agentes extraños, proteger al cuerpo de los rayos ultravioleta (UV) que causan daño a nivel del ADN y pueden derivar en el desarrollar cáncer de piel [41]; la siguiente capa, donde se encuentran los vasos sanguíneos, así cómo algunas glándulas y los folículos pilosos [41]; en la tercera y última capa de la piel, conocida cómo hipodermis, se encuentra el tejido adiposo y conjuntivo [41]. Por otro lado, hablando acerca de las células afectadas con el melanoma: los melanocitos, los cuales una vez maduros viajan desde la cresta neural (cuarto tejido embrionario [42]) hasta los folículos pilosos para comenzar a cumplir con su función principal: producir melanina (además de darle el pigmento a la piel, también ayuda a protegerla de la radiación UV [41].) y transportarla a los queratinocitos [14].



**Figura 4-1:** Capas de la piel humana. Disponible en [46]

Hablando acerca del melanoma cutáneo, este ha sido reconocido cómo una enfermedad por

más de dos mil años y su incidencia ha ido en aumento a nivel mundial desde que se comenzaron a recolectar datos en la década de los 60 [41]. Se desarrolla en los melanocitos [48], mostrando sus primeros síntomas apareciendo cómo un lunar nuevo o uno que cambio algún aspecto físico (tamaño, forma o color) [12] cómo se puede ver en la figura 4-2, y así cómo cualquier otro cáncer no existe una forma 100% segura para prevenirlo, sin embargo, los dermatólogos mencionan que existen varios factores de riesgo que aumentan la probabilidad de desarrollarlo [45]:



**Figura 4-2:** Ejemplo de un lunar con melanoma tomado de la base de datos ISIC. Disponible en [27]

- Exposición a radiación UV, especialmente en las horas en las que el índice UV se encuentra mayor a 7. Pues está comprobado que los rayos UVB al tener 1000 veces más potencia que los rayos UVA pueden causar daño directo al ADN pues favorecen la aparición de dímeros de pirimidina que generan mutaciones de C a T [45], por otro lado, los rayos UVA causan daño indirecto en el ADN al generar especies reactivas de oxígeno, que impiden la regeneración del ADN [45].
- Tener piel y cabello claro, en especial si hay historial de quemaduras solares graves o recurrentes, debido a que la cantidad de melanina en el cuerpo es menor, lo que deja a la piel más expuesta a los daños causados por la radiación UV [12]. Según el libro Melanoma del autor Riker, personas con el cabello rojo tienen un riesgo relativo de 3.64 de desarrollar este cáncer de piel, mientras que aquellas con ojos azules presentan un riesgo de 1.47 en comparación con las personas con ojos café que presentan un riesgo relativo de 2.06 [41].
- Tener historial médico de familiares con melanoma, aproximadamente el 10% de los

pacientes con melanoma cuentan con este antecedente [12], se estima que el tener un familiar en primer grado con melanoma puede aumentar hasta dos veces el riesgo[41].

- Personas con el sistema inmune deprimido [12].

El diagnóstico de melanoma en sus etapas tempranas puede reducir las tasas de mortalidad lo que como consecuencia trae una disminución en el costo invertido por el gobierno para el tratamiento y el cuidado de los pacientes [41], por esta razón se recomienda hacer una revisión periódica los lunares y en caso de notar alguna anomalía acudir al médico. Esta se puede hacer usando la Regla ABCDE, la cual fue diseñada por un grupo de investigadores de la universidad de Nueva York en 1985 como una guía fácil, rápida y sencilla de detectar alguna característica anormal en los lunares que permita la detección del melanoma en sus etapas tempranas, esta puede ser usada por doctores, dermatólogos en proceso de aprendizaje y público general [2]. Esta regla es un simple acrónimo que indica [6], para tener una mejor visualización de este ver figura 4-3:

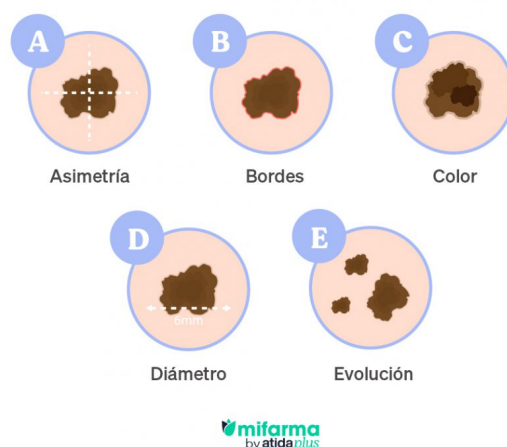
- **A: Asimetría**, si el lunar no tiene sus dos mitades similares.
- **B: Bordes**, si los bordes son irregulares o con terminaciones abruptas.
- **C: Colores**, si el lunar presenta más de un color.
- **D: Diámetro**, si el lunar tiene un diámetro mayor a 6 mm.
- **E: Evolución**, si el lunar presenta cambios en sus características físicas a lo largo del tiempo.

## 4.2 Base de datos ISIC Challenge

La base de datos empleada para este proyecto es la generada por la Sociedad Internacional de Imagen Digital de la Piel (*International Society for Digital Imaging of the Skin, ISDIS*), la cual a partir de 2016 lanzo la Colaboración Internacional para la Obtención de Imágenes de la Piel (*International Skin Imaging Collaboration, ISIC*) con una serie de retos que tenían como fin el mejorar el diagnóstico de melanoma. A lo largo de los años, el reto a escalado en cuanto a tamaño y complejidad, mientras que en un inicio se enfocaban en la clasificación y segmentación de melanoma para el ultimo reto lanzado, en 2020, las actividades se encontraban enfocadas en la falta de distribución y en evaluar el impacto que tienen estas tecnologías dentro de un ambiente clínico [27].

A la fecha, la base de datos es de acceso público que ser empleada para fines de enseñanza, investigación y desarrollo de algoritmos con base en métodos de inteligencia artificial [26],

## Regla ABCDE de los lunares



**Figura 4-3:** Regla ABCD para detectar melanoma. Disponible en [37]

tiene más de 13 mil imágenes dermatoscópicas provenientes de centros clínicos en todo el mundo, las cuales se encuentran, dependiendo del año en el que fueron añadidas al archivo, relacionadas con el diagnóstico dado por expertos, la segmentación, metadatos clínicos y características dermatoscópicas [27].

### 4.3 Inteligencia Artificial

La inteligencia artificial (IA) usa diferentes técnicas de análisis de datos como parte de la ciencia computacional para imitar la inteligencia humana y crear sistemas informáticos capaces de entender, aprender y demostrar, estas nacen como inspiración del aprendizaje humano donde a través de este se pueden detectar patrones, formas o texturas, tomar decisiones y manipular objetos [22], [36], su uso facilita el desarrollo de proyectos en áreas como la agricultura, la industria textil, la financiera y la medicina [33], como es el caso de este proyecto. Dentro de las ventajas que se encuentra el usar la IA dentro del contexto de la medicina se puede tener:

- Mejorar el cuidado de los pacientes y en consecuencia reducir los costos, al poder tener herramientas de apoyo diagnóstico que permitan la detección de enfermedades desde sus etapas más tempranas [4].
- Tratamientos más efectivos, al ayudar a los médicos a reducir la incertidumbre en la toma de decisiones de estos con base en los resultados de exámenes [38].

- Cirugías más precisas y exactas, con ayuda de la robótica y la IA, se reducen las complicaciones que se puedan presentar dentro de estos procedimientos [38].
- Facilitar interpretación de los datos clínicos, con ayuda del procesamiento de lenguaje natural se pueden analizar reportes de los pacientes y hasta realizar informes de exámenes [13].

### 4.3.1 Aprendizaje de máquina

El aprendizaje de máquina consiste en una serie de métodos o algoritmos encargados de detectar automáticamente patrones dentro de los datos y mejorar su desempeño con el tiempo [35]. En otras palabras, estos métodos usan su capacidad de aprendizaje y de adaptarse a los cambios junto con los datos que tienen para crear una aproximación del proceso, en el que están involucrados, con base en los patrones y con ello poder realizar predicciones [29].

A nivel más detallado, el aprendizaje de máquina se basa en el uso de diferentes teorías estadísticas que buscan optimizar un parámetro de rendimiento, es decir, modelos matemáticos que puedan predecir y/o describir [29]. En este caso, la informática toma un papel muy importante ante la creación de estos modelos: En primer lugar, la ejecución de algoritmos eficientes para entrenar, que cuenten con la capacidad de resolver el problema propuesto; y, en segundo lugar, asegurarse que esta eficiencia cubra todo el proceso posterior al entrenamiento [29].

El aprendizaje de máquina se divide en:

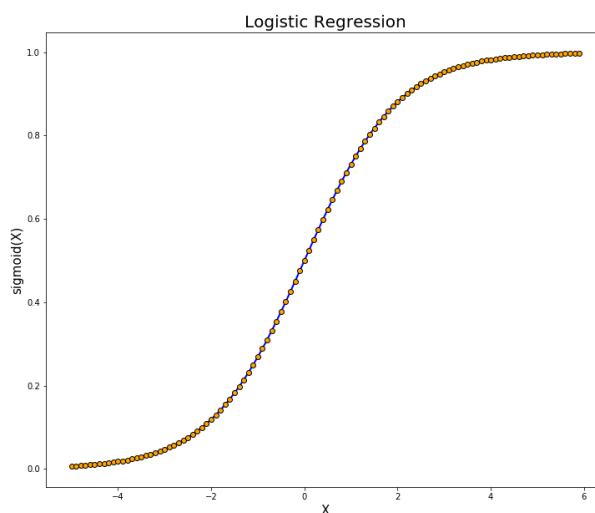
1. **Aprendizaje supervisado:** Los modelos basados en este tipo de aprendizaje relacionan las entradas con las salidas con base en el conjunto dado donde se encuentran relacionadas las entradas con las salidas [35].
2. **Aprendizaje no supervisado:** En este tipo, lo único que se le da al modelo son las entradas, con el fin de que este encuentre patrones dentro de los datos dados [35].
3. **Aprendizaje por refuerzo:** Es el tipo de aprendizaje de máquina menos usado, se basa en el aprendizaje del modelo dadas señales ocasionales de recompensa y castigo [35].

Adicionalmente, se define el aprendizaje profundo cómo aquel aprendizaje que usa redes neuronales convolucionales profundas para aprender de los datos ingresados [29], este se explicará con más detalle en secciones posteriores. A continuación, se explicarán los modelos empleados para realizar clasificación de las imágenes de melanoma con base en las características ABCD calculadas.

## Regresión logística

La regresión logística usa una función logística para determinar la probabilidad de que una entrada pertenezca a determinada clase, para ello transforma los valores de salida de una regresión lineal en una probabilidad que se encuentra entre 0 y 1. Debido a que es un modelo de aprendizaje de máquina supervisado y a las características mencionadas, este suele ser empleado en diversas tareas de clasificación [3].

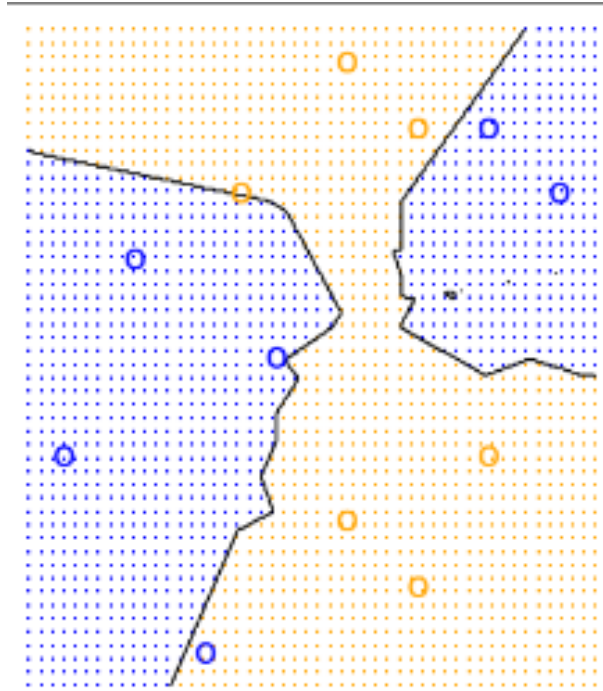
La función empleada en este modelo siempre va a dar una gráfica en forma de S, cómo la que se ve en la figura 4-4, por lo que sin importar el valor de entrada siempre se obtendrá una predicción razonable según los datos, así cómo, un mejor rango de probabilidades en comparación con una regresión lineal [3].



**Figura 4-4:** Gráfica representando la función empleada para regresión logística. Disponible en [51]

## K Nearest Neighbor

*K Nearest Neighbor* (KNN) es un modelo de aprendizaje supervisado, generalmente usado para clasificación, donde con base en el número  $K$  y el dato de entrada,  $X$ , predice los puntos  $K$  que se encuentran más cercanos a la entrada. Con ello se calcula la probabilidad condicional de que  $X$  pertenezca a cada uno de los puntos  $K$  más cercanos, y aquella con mayor valor será la clase de la entrada [3]. La figura 4-5 ilustra cómo funciona este modelo: Los puntos que se encuentren dentro de la zona azul serán de la misma clase, y lo mismo ocurre para los puntos dentro de la zona naranja [3].



**Figura 4-5:** Representación gráfica del modelo KNN. Disponible en[3]

### Random Forest

Este modelo, usado para clasificación y regresión (predecir valores continuos), usa una serie de árboles de decisión independientes, con la diferencia de que en cada división se elige una muestra aleatoria del conjunto total de indicadores, asegurándose que en ningún momento se escoja la mayoría de los indicadores, para al final combinar cada predicción y generar una predicción final. Esto se hace para que, en caso de tener un indicador con un peso considerable, este no sea escogido en la mayoría de las divisiones de los árboles generando una semejanza entre los árboles de decisiones; y cómo consecuencia un sesgo entre las predicciones de cada árbol. Es decir, el modelo *Random Forest*, tiene la ventaja de que el conjunto de predicciones no presenta una correlación muy fuerte, dando cómo consecuencia resultados menos variables y más confiables [3].

### Máquinas de soporte vectorial

Las máquinas de soporte vectorial (*Support Vector Machine, SVM*) son un modelo de aprendizaje de máquina en el que se busca un hiperplano que contenga la máxima distancia posible entre los datos de distintas clases, una vez obtenido este subespacio los datos más cercanos a este son seleccionados, vectores de soporte, y con ello se realiza el entrenamiento con el fin de que la distancia entre los vectores y el hiperplano sea la menor posible. Este método resulta muy útil en aquellos conjuntos de clases donde la frontera entre ellas no es lineal y

un clasificador de soporte vectorial no tendría un desempeño óptimo [3].

### Perceptron Multicapa

El perceptrón multicapa (*Multilayer perceptrón, MLP*) es un modelo de aprendizaje de máquina, que usa redes neuronales para clasificar o hacer regresión, mientras que el perceptrón por si solo define un hiperplano y la red neuronal perceptrón lo implementa, estas son útiles a la hora de aproximar funciones lineales, pero no resuelven tan bien la aproximación de funciones no lineales, esta es tarea del perceptrón multicapa. El cual consiste en una serie de capas conectadas entre si en las cuales se tiene el conjunto de datos de entrada, el MLP se encarga de calcular con una función de activación la salida de cada una de las neuronas de cada capa y pasar esa información a las siguiente, para que la salida se pueda usar para predecir el valor de salida correspondiente [29].

### 4.3.2 Aprendizaje profundo

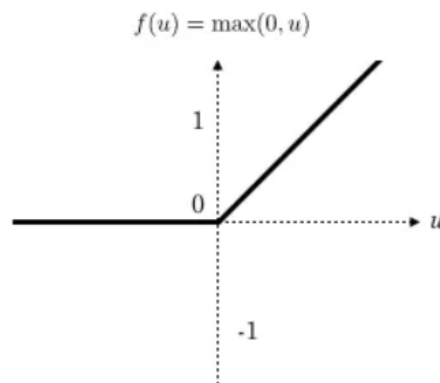
El aprendizaje profundo usa las redes neuronales para aprender de la información brindada por los datos, donde a partir de la entrada, cada capa oculta combina los valores dados por la capa anterior y los propios para realizar transformaciones diferentes hasta llevar a un nivel abstracto, que permita llegar a la salida representada con estas definiciones. La idea de llegar a ese nivel abstracto, se basa en que para la mayoría de las tareas no se cuenta con el conocimiento acerca de la estructura que hay en la entrada y, por lo tanto, cualquier patrón o característica que exista debe ser aprendida durante el entrenamiento [30], [29]. En otras palabras, estos métodos resultan muy llamativos pues no necesitan de la mediación del ser humano para extraer características, pues para ello usan una jerarquización de conceptos que permite obtener características más complejas y con mayor utilidad a la hora de hacer las diferentes tareas bien sean de clasificación, regresión, generación o recuperación de datos [21], [30], [29].

El aprendizaje profundo ha sido de mucha utilidad para descubrir patrones o estructuras complejas en grandes conjuntos de datos, lo que lo hace aún más llamativo para usar en reconocimiento de imágenes y de habla, cómo apoyo en áreas cómo la farmacología, la química y la medicina [30]. De manera más específica, existen un tipo de redes neuronales [47], usadas en este proyecto, diseñadas para procesar imágenes de entrada: Las redes neuronales convolucionales (CNN). Estas realizan la extracción de características automáticamente, usando una diversidad de capas cómo las expuestas en detalle a continuación [47]:

1. **Capa convolucional:** La capa principal de las CNN, se encarga de detectar conjuntos de características de las imágenes. Para ello hace un filtro convolucional en el que usa un *kernel* (característica) para hacer la convolución con la imagen y obtener el valor de esta operación. Esto arroja a la salida un mapa de características, el cual indica

el valor de similitud entre la imagen y la característica, es decir entre mayor sea este, mayor semejanza se tiene y viceversa [47].

2. **Capa de *pooling*:** Generalmente se encuentra entre dos capas convolucionales, se encarga de reducir el tamaño de las imágenes que recibe de entrada manteniendo las características relevantes. Para ello, se corta la imagen en porciones regulares y se mantiene el valor máximo dentro de cada una de ellas, obteniendo en la salida el mismo número de mapas de características de la entrada, con la diferencia que estos son más pequeños, esto tiene la ventaja de mejorar la eficiencia de la CNN y evita el sobre aprendizaje [47].
3. **Capa de corrección ReLU (*Rectified Linear Units*):** Usa la función real no lineal, ver gráfico de la función en la figura 4-6, para reemplazar todos los valores negativos en la imagen de entrada por ceros [47].
4. **Capa densa:** Siempre se encuentra en la ultima parte de las redes neuronales, recibe un vector en la entrada, el cual es multiplicado por una matriz de pesos para obtener a la salida un vector nuevo, del tamaño de las clases, que indica la posibilidad de que la imagen pertenezca a cada una de estas [47].



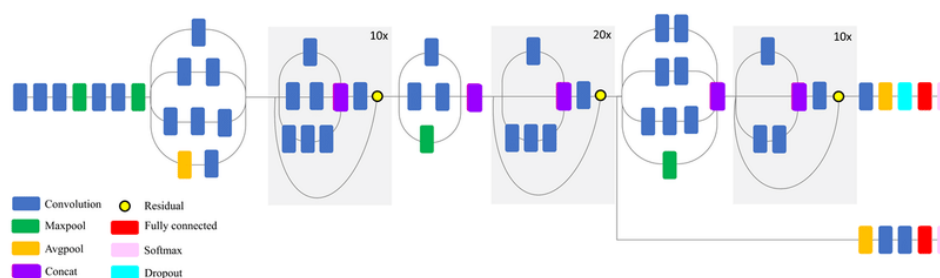
**Figura 4-6:** Gráfica de la función ReLU. Disponible en[47]

La forma y el orden en el que se conectan estas capas genera distintas de arquitecturas, que pueden tener mejores resultados según el problema de investigación en el que se emplean. Para este proyecto se empleados dos arquitecturas diferentes, *InceptionResNet* y *EfficientNet*, explicadas a continuación.

### InceptionResNet

La arquitectura *InceptionResNet*, ver figura 4-7, une dos de las redes neuronales convolucionales (CNN) que han presentado mejores resultados generales, *ResNet* e *Inception*. Esta

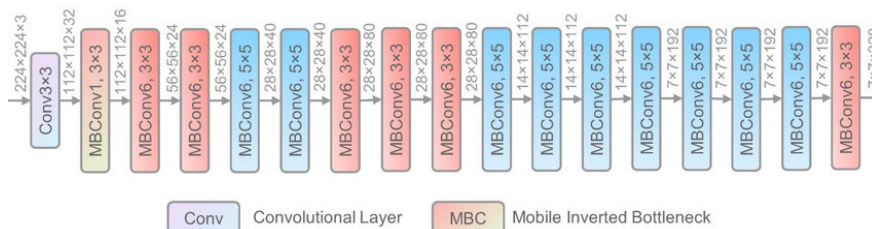
arquitectura cuenta con un total de 164 capas en las cuales la normalización de lotes se usa únicamente en las capas tradicionales, adicionalmente se usan un total de residuales que permiten usar un mayor número de capas de la arquitectura *Inception*. En términos generales esta CNN combina varios bloques de las dos CNN mencionadas anteriormente, toma como entrada imágenes a color bien sea de  $224 \times 224 \times 3$  o  $299 \times 299 \times 3$  [31], [16].



**Figura 4-7:** Representación gráfica de la arquitectura InceptionResNet. Disponible en[31]

## EfficientNet

La arquitectura *EfficientNet*, ver figura 4-8 es una red neuronal convolucional (CNN) lanzada por *Google*, que usa un método para escalar de manera uniforme todas las dimensiones de la imagen con base en un coeficiente compuesto, esto se hace con la idea de que, para imágenes de grandes dimensiones, la red empleada para entrenar requerirá de mayor cantidad de capas y de canales para lograr extraer las características y los patrones más pequeños. Este escalamiento permite tener una mejor precisión sin aumentar la cantidad de recursos requeridos durante el entrenamiento [49], [53].



**Figura 4-8:** Representación gráfica de la arquitectura EfficientNet. Disponible en[53]

# 5 METODOLOGÍA

En la figura 5-1 se encuentra las fases ejecutadas durante el desarrollo de este proyecto, cada una asociada al cumplimiento de un objetivo específico. A continuación se explica de manera detallada cada una de las etapas.

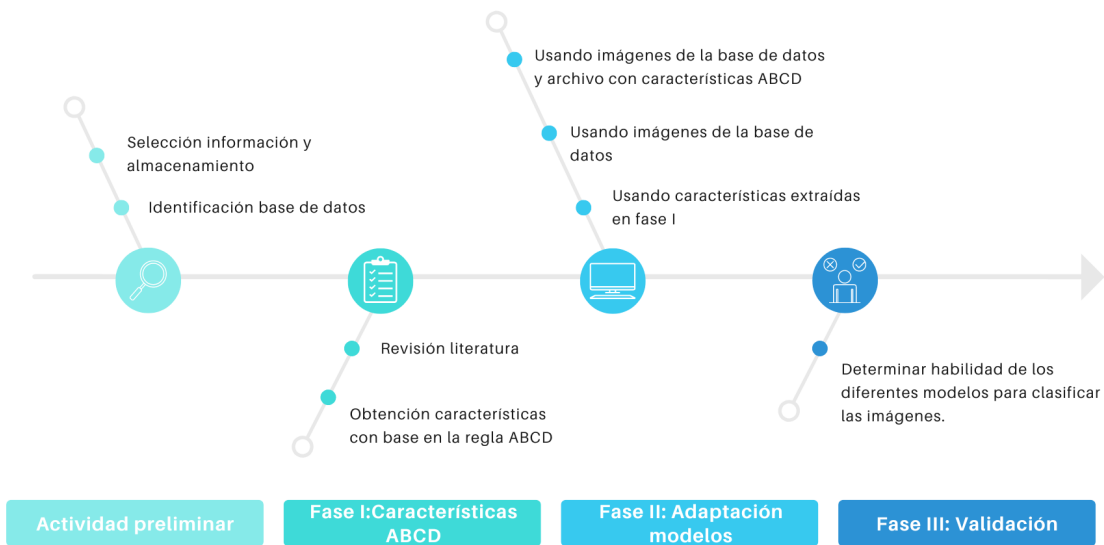


Figura 5-1: Diagrama general de metodología

## 5.1 Actividad preliminar: Identificación de la base de datos, selección de la información y almacenamiento

Durante esta actividad preliminar se realizó la identificación de la base de datos con el fin de comprender que información la conformaba, que tan confiable era, las autorizaciones había que tener en caso de necesitarlas o si, por el contrario, era una base de datos abierta. Una vez realizado este reconocimiento, se entiende que esta es una base de datos abierta patrocinada

por la sociedad internacional de la imagen digital de la piel (ISDIS) que puede ser empleada para investigación o educación, [26] conformada por imágenes dermatoscópicas tomadas de clínicas y centros dermatológicos ubicados en todo el mundo [27].

Posterior a esto se realiza la selección de la información que resulta útil para la realización de este proyecto, esta se hace con base en los diferentes propósitos de cada uno de los retos lanzados por la organización, entendiendo que cada año la complejidad de estos aumentaba con el fin de diagnosticar más lesiones de la piel diferentes a melanoma, y teniendo en cuenta el objetivo general de esta investigación. Se seleccionaron aquellas imágenes correspondientes a los retos enfocados en detectar melanoma en lunares, aquellos que fueron lanzados en los años 2016 y 2017, este grupo de imágenes se encuentra relacionado con la imagen a color, la imagen con el lunar segmentado a blanco y negro y el diagnóstico dado por expertos.

Luego de esta selección, con las 4029 imágenes, se procedió a hacer el almacenamiento, Pero antes de eso, se debió hacer un redimensionamiento de las imágenes para que todas quedaran con el mismo tamaño, 224 x 224 pixeles, esto se hizo debido a que muchas arquitecturas de aprendizaje profundo requieren de un tamaño específico y que este sea el mismo para todas las imágenes de entrada. Posterior a esto se realizó el almacenamiento de las imágenes en *Google Drive* divididas en las carpetas correspondientes de entrenamiento y prueba, siendo esta una proporción de 70% para entrenamiento y 30% para prueba tal como viene seleccionado desde la base de datos, para poder ser usadas en las siguientes fases usando *Google Colaboratory*.

## 5.2 Fase I: Obtención de las características de los lunares con base en la regla ABCD.

Para la obtención de características correspondientes a la regla ABCD de los lunares se hizo una revisión de literatura usando *Google Scholar* y los recursos electrónicos de la Universidad del Rosario de artículos científicos con un tiempo de publicación no mayor a 10 años con palabras clave: melanoma, regla ABCD y extracción de características con el fin de recolectar información acerca de cómo han sido los acercamientos de estudios previos a la hora de calcular esta regla usando técnicas clásicas de procesamiento de imágenes. Con esto se encontró la revisión realizada por Ali et al. en el que evalúan cómo diferentes estudios han automatizado la regla ABCD para melanoma [2]. A continuación, se expondrá como se realizó la obtención de cada una de las cuatro características: Asimetría, Bordes Irregulares, Colores, Diámetro, las cuales al finalizar esta fase se introdujeron en un archivo donde se encuentran relacionados estos valores, con las imágenes y su diagnóstico correspondiente.

### 5.2.1 Asimetría

Para hallar la asimetría de cada imagen, se determinó que es importante hallarla en el eje horizontal y en el vertical, por lo cual para esta característica se obtienen dos valores por cada una de las imágenes. Esto se hace usando las máscaras de las imágenes de la base de datos, debido a que la información a obtener de la imagen depende enteramente de su forma y no de sus colores, por lo que trabajar usando la imagen binaria facilita este procesamiento.



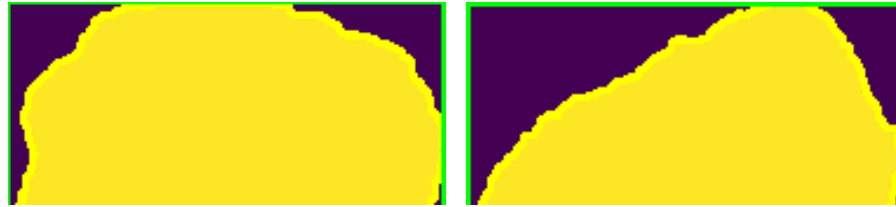
**Figura 5-2:** Máscara del lunar con su correspondiente bounding box

En primer lugar, se halló el *bounding box* de la imagen, esto se hizo encontrando los contornos de esta y graficando el rectángulo correspondiente a la región de interés, figura 5-2, sobre la cual se deben hallar las dos asimetrías. Posterior a esto se comenzó a trabajar con los pixeles que se encuentran al interior del *bounding box*, ubicando, en primer lugar, la fila y la columna que se encuentran justo en el medio de esta región de interés, y con esto se realizó la división de la imagen, es decir, las dos mitades (verticales y horizontales) del lunar. Luego, se dio la vuelta a una de las dos mitades para que todas quedaran apuntando a la misma dirección, como se ve en la figura 5-3.

Y con estas mitades, se usa el índice *Jaccard* para determinar el nivel de similitud entre las dos, esta métrica suele usarse para tareas de segmentación en las cuales se desea mirar el nivel de semejanza entre las máscaras obtenidas por los modelos y las del experto. De esta forma entre más cercanos a 1 sean los valores, más similar son las mitades, es decir, los lunares son más simétricos.

### 5.2.2 Bordos irregulares

Para determinar si los bordes son irregulares o no, se usó la formula expuesta en el estudio publicado por Golston et al. [20] mencionados en la revisión realizada por Ali et al. [2]. En la



(a) Mitad superior de la región de interés de la figura 5-2 (b) Mitad inferior invertida de la región de interés de la figura 5-2



(c) Mitad izquierda de la región de interés de la figura 5-2 (d) Mitad derecha de la región de interés de la figura 5-2

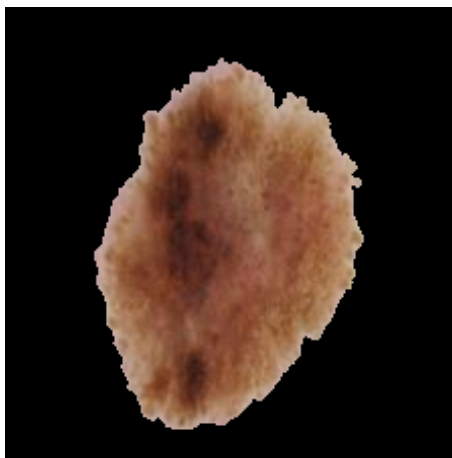
**Figura 5-3:** Las cuatro mitades halladas para encontrar el valor de asimetría

cual se halla la irregularidad usando la formula:  $I = \frac{P^2}{4\pi A}$  donde  $P$  corresponde al perímetro de la imagen y  $A$  al área, es decir que, los valores más altos de irregularidad indican bordes más regulares.

Para las imágenes de la base de datos, se usaron las máscaras de los lunares, a ellas se les halló el contorno de la región de interés, y se calculó el perímetro y el área de este; una vez obtenidos estos valores se emplea la formula mostrada anteriormente, para hallar la irregularidad de los bordes. Es decir, que para esta característica se halló un valor único asociado a cada una de las imágenes.

### 5.2.3 Colores

Para esta característica se usaron las imágenes a color y las máscaras de estas, esto con el fin de multiplicar ambas imágenes y tener una imagen donde el fondo es de color negro y el lunar es a color, como se ve en la figura 5-4. Con esta imagen se obtiene el valor de cada uno de los pixeles del lunar para cada uno de los tres canales, teniendo en cuenta que el fondo al ser negro tiene un valor de 0 y se le asignó un valor NaN para que no sume a la hora de determinar la cantidad de colores. Finalmente, se determina la cantidad de colores que hay en la imagen para cada uno de los canales, es decir, de esta característica se obtiene tres valores, uno por los colores en cada canal (rojo, verde y azul).



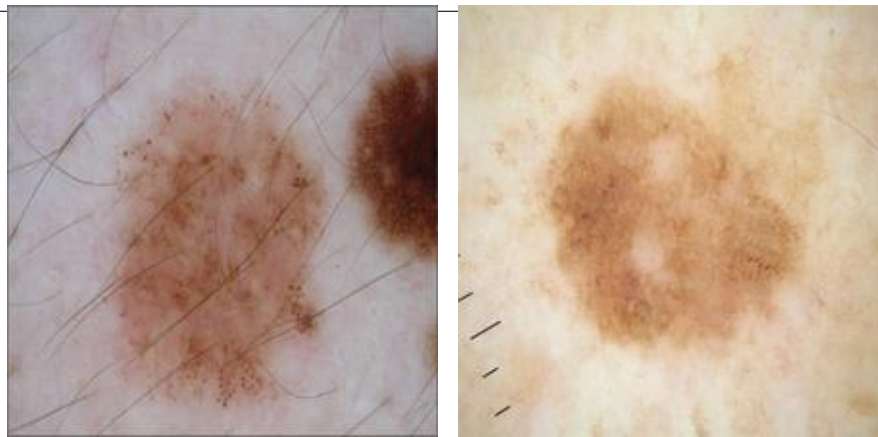
**Figura 5-4:** Imagen obtenida de la multiplicación de la imagen a color y su respectiva máscara

#### 5.2.4 Diámetro

En diámetro, se usó únicamente la máscara del lunar a blanco y negro, con esta se halló el contorno de la región de interés para posteriormente hallar sus coordenadas, es decir las coordenadas de lo que corresponde al *bounding box* que rodea únicamente el área rectangular que abarca el lunar. Con estas coordenadas se calculó la distancia entre cada uno de los puntos, es decir, todos los diámetros; y con este conjunto de diámetros se buscó aquel con mayor valor para asignarle a este el valor de esta característica. Cabe resaltar, que como las imágenes no cuentan con un punto de comparación para poder hacer la transformación de píxeles a milímetros, esta característica quedo de esta forma sin poder determinar si un lunar tenía un diámetro mayor a 6mm como lo indica la regla ABCD.

### 5.3 Fase II: Adaptación de modelos de aprendizaje profundo para la extracción de características y clasificación de melanoma de las imágenes dermatoscópicas de lunares

Como se ha mencionado anteriormente, la clasificación de melanoma no es una tarea sencilla para aquellos médicos que no tienen la experiencia a la hora de detectar estas lesiones, y si bien el uso de inteligencia artificial puede ayudar a esta tarea, esta tampoco resulta ser una tarea sencilla para la tecnología. Pues debido a las características de la piel como el tono, el color, los vellos, las cicatrices y hasta los lunares, sin importar si estos son benignos o malignos, pueden resultar muy similares entre sí, como se puede ver en la figura 5-5, lo que dificulta la tarea de estos modelos para clasificar de manera acertada esta enfermedad;



(a) Lunar sin melanoma

(b) Lunar con melanoma

**Figura 5-5:** Comparación lunar benigno y uno con melanoma. Imágenes disponibles en [26].

por este motivo, se propuso el mezclar la extracción de características que hacen las CNN normalmente, junto con las características extraídas manualmente de la regla ABCD para mejorar el desempeño de los modelos para clasificar melanoma.

### 5.3.1 Adaptación de los modelos para clasificar melanoma usando las características ABCD extraídas en la Fase I

En esta etapa, se usaron modelos de aprendizaje de máquina clásicos y se hizo una exploración de algunos otros predeterminados usando la librería *Lazy Predict* para realizar la clasificación de melanoma basada únicamente en las características extraídas de la regla ABCD de la fase I. Para la primera parte se usaron 5 clasificadores distintos:

1. **Regresión logística:** Con este clasificador se hizo un *Grid Search* para explorar el valor del parámetro C de este modelo (Inverso de la fuerza de regularización) de valores entre 1 y 1000000 en escala logarítmica. Esta primera búsqueda arrojó que  $C = 100000$  es el mejor parámetro, y con este se realizó la predicción del modelo para observar su desempeño en entrenamiento y en prueba usando métricas como la matriz de confusión, sensibilidad, exactitud balanceada, especificidad y coeficiente *Kappa*.
2. ***K-nearest neighbors*:** Con este clasificador se estableció el valor de  $K = 1$ , para observar su comportamiento con este conjunto de datos, luego del entrenamiento se analizaron las mismas métricas de desempeño usadas en regresión logística, tanto en prueba como en entrenamiento y con esto validar la habilidad del modelo.
3. **Ranfom Forest:** : Al igual que con regresión logística se realizó un *Grid Search* para realizar exploración del parámetro de estimadores con valores entre 5 y 100, y saber cuál

| Modelos             | Parámetros explorados    |                                  |                                  |                       |
|---------------------|--------------------------|----------------------------------|----------------------------------|-----------------------|
| NearestCentroid     | <i>Metric</i> : Todos    | <i>Shrink threshold</i> : [0, 1] |                                  |                       |
| ExtraTreeClassifier | <i>Criterion</i> : Todos | <i>Splitter</i> : Todos          | <i>Max Depth</i> : 2, 13, 18, 20 |                       |
| LabelPropagation    | <i>Kernel</i> :Knn y Rbf | <i>Gamma</i> : 0.3, 0.7, 1       | <i>N Neighbors</i> : 3,5,7       |                       |
| LabelSpreading      | <i>Kernel</i> :Knn y Rbf | <i>Gamma</i> : 0.3, 0.7, 1       | textitN Neighbors: 3,5,7         | <i>Alpha</i> : [0, 1] |

**Tabla 5-1:** Parámetros empleados explorar usando GridSearch

de estos puede resultar el mejor para el entrenamiento, siendo este el correspondiente a 100 estimadores. Una vez realizado esta exploración de ajuste fino se realiza el respectivo entrenamiento y prueba del modelo para luego evaluar su desempeño con las métricas mencionadas anteriormente.

4. **Máquinas de soporte vectorial:** Se empleó la función *Grid Search* para realizar la exploración de todas las combinaciones posibles de los parámetros  $C$  (parámetro de regularización) y el coeficiente *kernel*, gamma ubicado en valores entre 0 y 64. Esta exploración arroja que  $C$  y gamma igual a 0.03125 dan el mejor rendimiento, sobre estos parámetros es que se hizo el entrenamiento y prueba del clasificador de máquinas de soporte vectorial sobre el cual se evaluará la habilidad del modelo con base en los valores de las métricas de desempeño.
5. **Perceptrón multicapa:** Con este clasificador se modificó únicamente el número de neuronas y de capas ocultas ( $64 \times 256 - 256 \times 512 - 512 \times 1024$ ) y se realizó el mismo procedimiento que con los anteriores clasificadores: se observó el entrenamiento y prueba del modelo con las métricas de desempeño seleccionadas para evaluar la habilidad del modelo a la hora de clasificar melanoma usando los datos de las características ABCD.

De manera adicional, se usó la librería *Lazy Predict* con el objetivo de tener una idea general acerca de la evaluación de diversos modelos de aprendizaje de máquina y poder saber cuáles presentan unas mejores métricas, con base en la tabla de modelos que proporciona esta librería. Su predicción mostro 26 modelos con una exactitud balanceada entre 50% y 68%, de estos 26 modelos se usaron los 4 mejores (*NearestCentroid*, *ExtraTreeClassifier*, *LabelPropagation*, *LabelSpreading*) para emplear *Grid Search* y conocer bajo cuales parámetros, ver tabla 5-1, se tenía el mejor desempeño.

Esta combinación de parámetros arroja aquellas combinaciones que tenían un mejor rendimiento, ver tabla 5-2, con las cuales se realizó el entrenamiento y prueba de los modelos usando las características ABCD, y a partir de esto validar los modelos analizando los valores de las métricas de desempeño empleadas.

| Modelos             | Combinación  |
|---------------------|--|
| NearestCentroid     | <i>Metric: Cityblock Shrink threshold: 0.2</i>       |
| ExtraTreeClassifier | <i>Criterion: Gini Splitter: Random Max Depth: 2</i> |
| LabelPropagation    | <i>Kernel:Knn Gamma: 0.3 N Neighbors:7</i>           |
| LabelSpreading      | <i>Kernel:Knn Gamma: 0.3 N Neighbors:7 Alpha:0.9</i> |

**Tabla 5-2:** Mejores combinaciones como resultado de emplear GridSearch

### 5.3.2 Adaptación de los modelos para clasificar melanoma usando las imágenes de la base de datos

Para la adaptación de los modelos empleados para clasificar melanoma se usó la arquitectura *InceptionResNetv2*, antes de hacer las adaptaciones necesarias de los modelos, se usó *ImageDataGenerator* para realizar aumento de datos, a través de rotaciones, acercamientos, desplazamientos horizontales y verticales de las imágenes; esto buscando mejorar el rendimiento de los modelos. Posterior a esto, se hacen las adaptaciones de los modelos, siendo estas: un congelamiento inicial de los pesos, para luego agregar cuatro capas adicionales (*GlobalAveragePooling2D*, *BatchNormalization*, *Dropout* y *Dense*).

Una vez realizadas estas modificaciones se inicia con el entrenamiento, el cual se realizó en tres etapas:

1. En la primera etapa se congelaron los pesos de las capas previo al entrenamiento, el cual se hizo con un tamaño de lote de 128, 10 épocas y usando para los datos de validación el conjunto de prueba.
2. En esta segunda etapa se descongelaron los pesos de las capas, se realizó el entrenamiento cargando los pesos del modelo de la primera etapa junto con un tamaño de lote de 32, 30 épocas y usando nuevamente los datos de prueba como validación.
3. En la tercera y última etapa, muy similar a la segunda, se descongelaron los pesos de las capas, se realizó el entrenamiento luego de cargar los pesos del modelo obtenido de la segunda etapa con un tamaño de lote de 32 y 20 épocas.

### 5.3.3 Adaptación de los modelos para clasificar melanoma usando las imágenes de la base de datos y el archivo de características ABCD

En esta etapa se emplearon dos arquitecturas diferentes *InceptionResNetV2* y *EfficientNetV2L*, usando *late fusion*: en donde se extraen las características y se combinan con las obtenidas durante la Fase I, para mejorar el desempeño del modelo a la hora de clasificar,

esta combinación se realiza al final del modelo en las capas densas. De manera adicional se agregaron 3 capas al modelo, una capa de *GlobalAveragePooling2D* que luego se concatena con un vector de 7 elementos (correspondiente a los valores obtenidos de la regla ABCD en la Fase I); una capa *BatchNormalization* y una *Dropout*. A partir de este momento se realizan tres entrenamientos diferentes para observar como el adicionar cada una de las capas puede afectar el rendimiento del modelo, es decir, para el primer entrenamiento se usó únicamente la capa concatenada adicional; para el segundo se adicionó la capa concatenada y *BatchNormalization* y para el tercero las tres capas adicionales mencionadas anteriormente.

En todos los entrenamientos realizados, usando las dos arquitecturas diferentes, se tuvieron 30 épocas, se usaron los datos de prueba como validación y se modificó el tamaño del lote entre 128 y 64, esto debido al espacio de memoria con el que cuenta la plataforma de *Google Colaboratory*, Adicionalmente se usaron dos monitores: Parada anticipada, para que se detenga el entrenamiento cuando pasadas 14 épocas la exactitud no presente mejoras; y *Model Checkpoint callback*, que permite guardar los mejores pesos del modelo con base en el menor valor de pérdida en validación obtenido durante el entrenamiento. Con estos modelos se obtuvieron métricas de desempeño como matriz de confusión, exactitud balanceada, especificidad, sensibilidad, entre otras, las cuales fueron empleadas en la fase IV para hacer la validación de todos los modelos y determinar la habilidad que poseen para clasificar melanoma.

## 5.4 Fase III: Validación de la habilidad del método para clasificar melanoma usando las imágenes de la base de datos ISIC.

En la última fase, usando las métricas de desempeño de cada uno de los modelos en entrenamiento y prueba se determina la habilidad de estos para realizar la tarea de clasificación de las imágenes y poder hacer una comparación con los antecedentes mostrados en la sección 2.1 y entre cada uno de ellos para determinar cuál de los tres conjuntos de experimentos mostraron mejor desempeño general y clasificando melanoma.

Es importante mencionar que cada una de las métricas empleadas permiten tener una idea diferente acerca de la calidad de los modelos, por ejemplo:

- **Exactitud:** Es la relación de predicciones correctas totales que tuvo el modelo dividido por el total de datos [23].
- **Especificidad:** Es la relación de las predicciones negativas correctas sobre el total de predicciones negativas del modelo [23].

| Valor       | Interpretación |
|-------------|----------------|
| Menor que 0 | Pobre          |
| 0 - 0.2     | Ligero         |
| 0.2 - 0.4   | Justo          |
| 0.4 - 0.6   | Moderado       |
| 0.6 - 0.8   | Considerable   |
| 0.8 - 1     | Casi Perfecto  |

**Tabla 5-3:** Interpretación del coeficiente Kappa. Disponible en [44]

- **Sensibilidad:** Es la relación de las predicciones positivas correctas sobre el total de predicciones positivas del modelo [23].
- **Coefficiente *kappa*:** Su valor indica el nivel de acuerdo entre el experto y la predicción realizada por el modelo, como se ve en la tabla **5-3** [44]

# 6 RESULTADOS

A continuación, se muestran los resultados obtenidos durante la realización de esta investigación, cada uno de estos asociados a las fases I a la III de la metodología.

## 6.1 Resultados de la obtención de las características ABCD

Una vez calculadas las características de asimetría, bordes, cantidad de colores y diámetro de todas las imágenes, estas fueron asociadas al nombre de la imagen y su etiqueta de melanoma (1) o benigno (0) y se guardaron en un archivo tipo CSV. Con esto se realizó un análisis estadístico para determinar si existen diferencias entre los valores de cada característica para las dos clases, específicamente, se graficaron diagramas de cajas para las 7 características (asimetría horizontal, asimetría vertical, bordes irregulares, cantidad de colores en el canal rojo, verde, azul y diámetro) para observar la distribución de los datos, complemento a esto se realizó un gráfico de dispersión para comprender la dispersión de los datos, junto con un gráfico de densidad para todos los pares posibles de características, esto permitió observar si existe una diferencia significativa en la distribución de los dos pares de características observadas comparándolas con las dos clases de la base de datos.

Con esto en primer lugar para asimetría, se observó que la media de los datos con diagnóstico benigno tenía un mayor valor, correspondiente a 0.849 en comparación con la media de 0.816 de las imágenes malignas, como se puede apreciar en la figura **6-1** la cual corresponde a la asimetría horizontal de las imágenes de entrenamiento, lo que indica que aquellas con lunares benignos presentan mayor simetría que las imágenes de lunares con melanoma, lo mismo ocurre con las imágenes de prueba, como se ve en la figura **6-2** donde aquellas que tienen una etiqueta de diagnóstico 0 tienen un mayor valor para la asimetría horizontal.

Por otro lado, la asimetría vertical presenta un comportamiento similar al observado con los diagramas de cajas de la asimetría horizontal, los grupos tanto de entrenamiento como de prueba presentan una media mayor para el conjunto de imágenes asociadas a un diagnóstico benigno, 0.884 en comparación con 0.842, como se puede ver las figuras **6-3** y **6-4**.

En el caso de los bordes, también se presentaron bordes más regulares para el conjunto de imágenes con etiqueta de diagnóstico 0, ver figura **6-5**, sin embargo, si se presentó una mayor cantidad de datos extremos que pudo afectar el diagrama y la media, pues para el conjunto

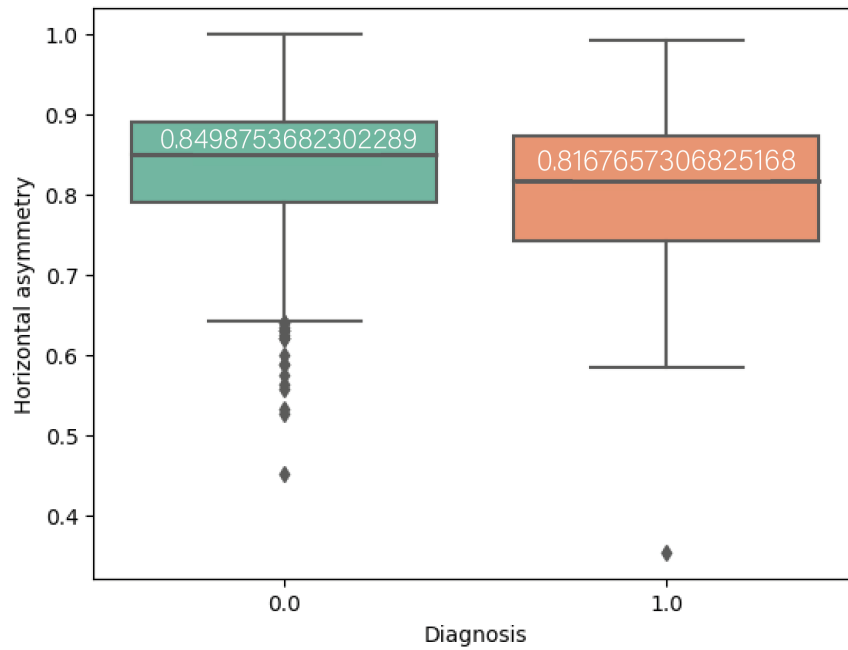


Figura 6-1: Diagrama de cajas del conjunto de entrenamiento para la característica de asimetría horizontal

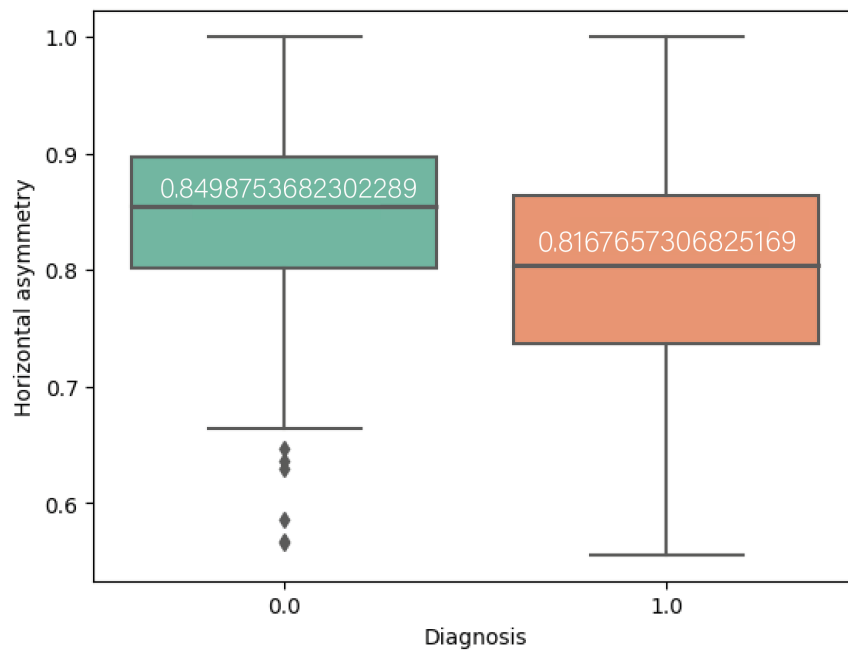
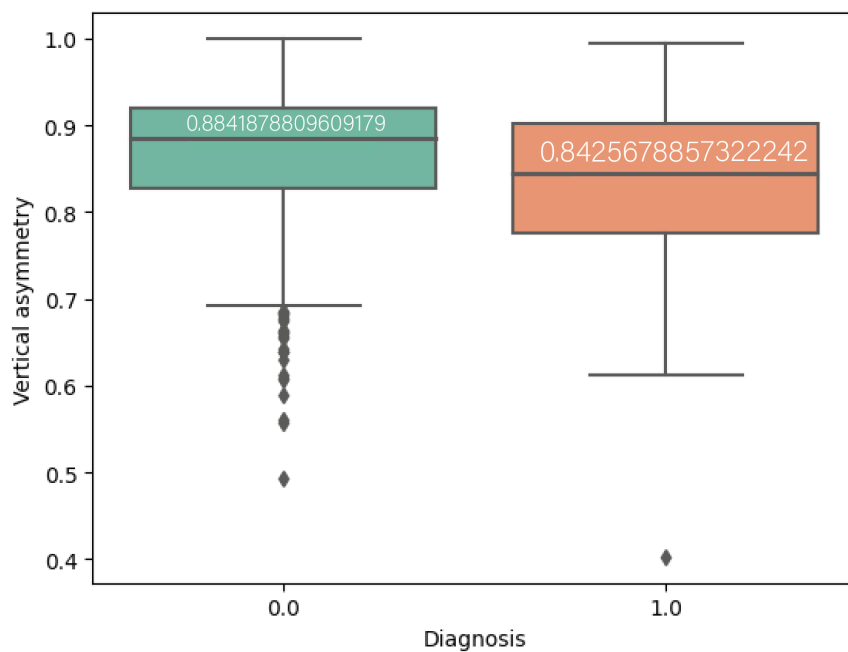
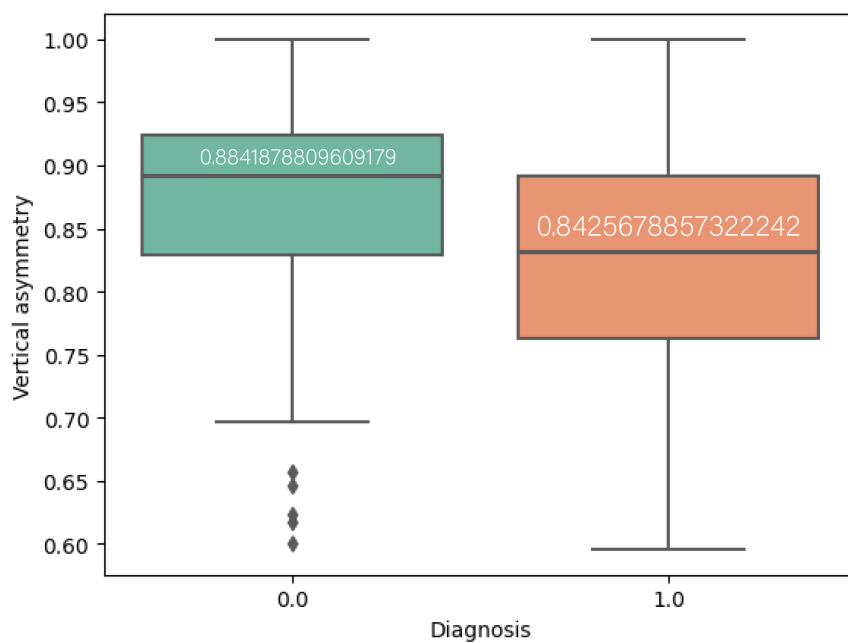


Figura 6-2: Diagrama de cajas del conjunto de prueba para la característica de asimetría horizontal



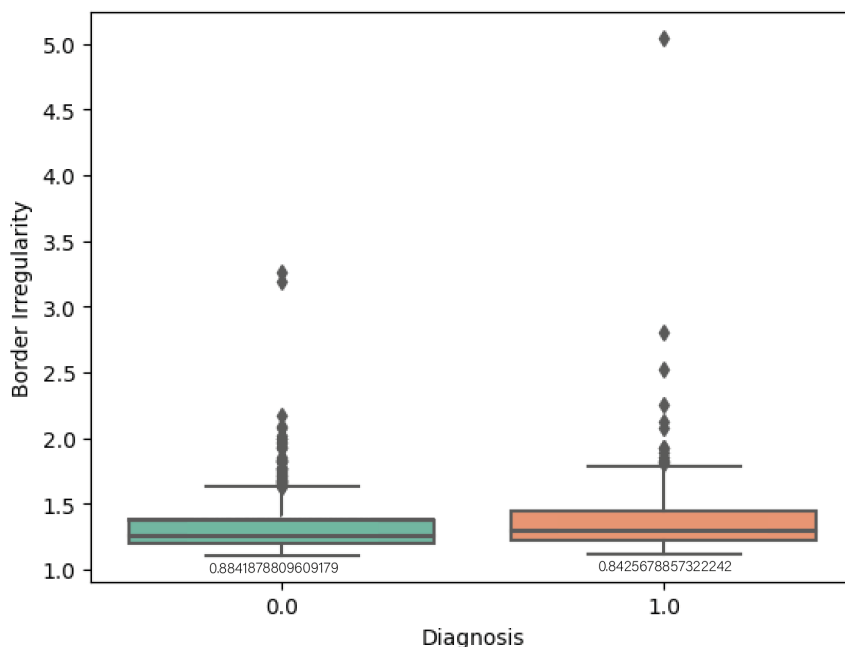
**Figura 6-3:** Diagrama de cajas del conjunto de entrenamiento para la característica de asimetría vertical



**Figura 6-4:** Diagrama de cajas del conjunto de prueba para la característica de asimetría vertical

de prueba, se tuvieron medias superiores a 1, 1.259 para las imágenes benignas y 1.286 para

las imágenes malignas, pero observando la revisión realizada en 2020 por Ali et al, [2] se menciona que aquellos valores superiores a 1.8 son imágenes muy irregulares; es decir, que para este conjunto, las imágenes presentan bordes aún más irregulares que las imágenes de entrenamiento.

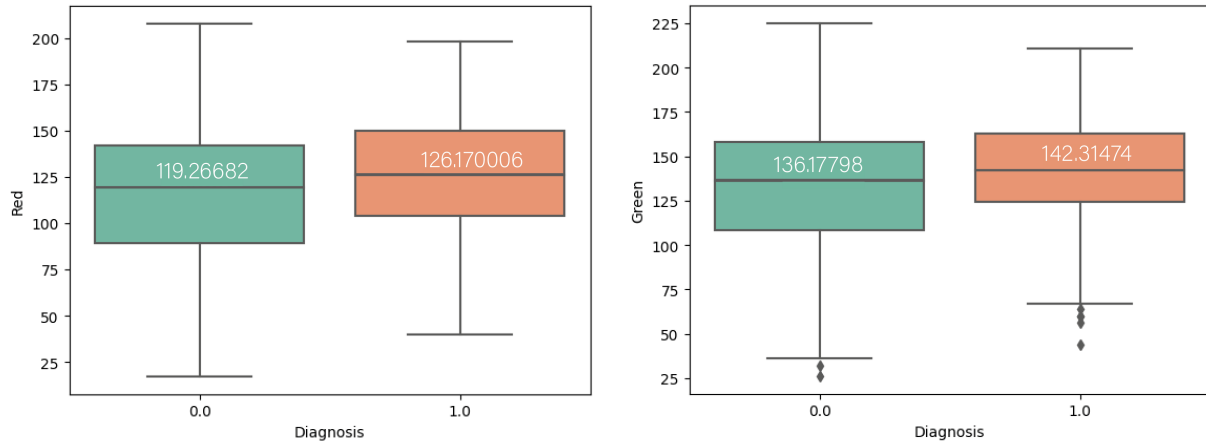


**Figura 6-5:** Diagrama de cajas del conjunto de entrenamiento para la característica de bordes

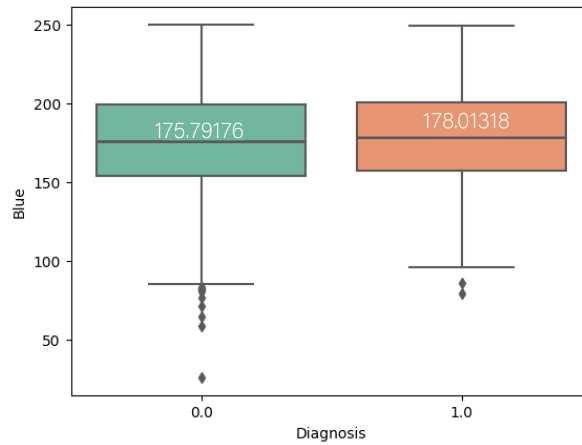
Con la cantidad de colores se tuvo que, para todos los canales del conjunto de entrenamiento y prueba, las imágenes con etiqueta 0 tenían una menor cantidad de colores que aquellas con etiqueta 1, como se ve en las figuras 6-6. Esto indica que, así como la dice la regla ABCD, los lunares con melanoma presentan una mayor cantidad de colores, los cuales pueden tomar distintas tonalidades de café, negro e incluso azules, donde para este canal se tuvo la mayor cantidad de colores, una media de 175.791 para las etiquetas 0 y 178.013 para las etiquetas 1, este comportamiento fue muy similar para el conjunto de entrenamiento y prueba.

Con el diámetro, fue donde se halló una diferencia más pronunciada entre las medias de las imágenes con lesiones benignas y las malignas, 138.924 y 183.991 respectivamente, esto para el conjunto de entrenamiento como se puede ver en la figura 6-7, para prueba la media tuvo un comportamiento similar al ya mencionado.

Con la gráfica de dispersión, se observó un comportamiento muy similar entre las imágenes benignas y malignas para todas las características, notando que debido a que hay una mayor



(a) Diagrama de cajas de colores en el canal rojo (b) Diagrama de cajas de colores en el canal verde

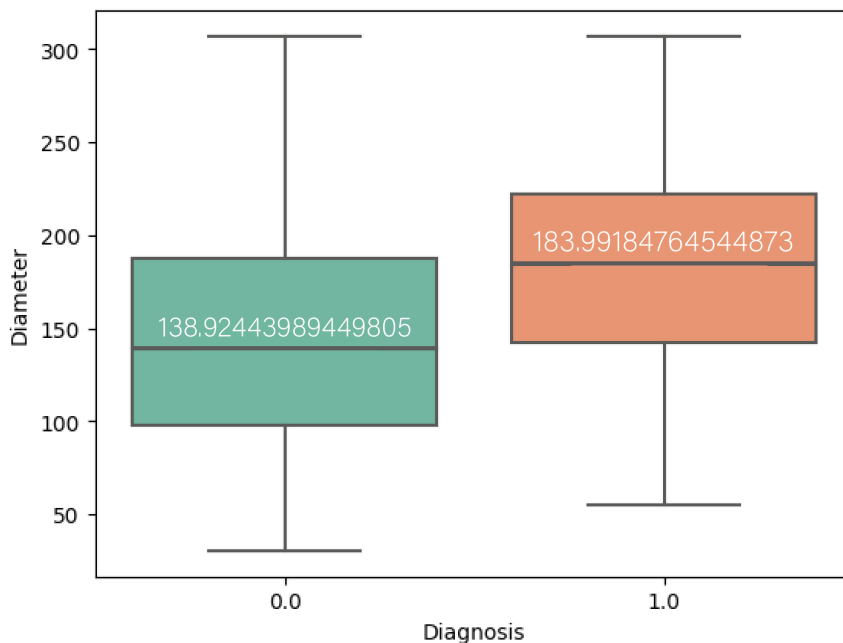


(c) Diagrama de cajas de colores en el canal azul

**Figura 6-6:** Diagrama de cajas del conjunto de entrenamiento para la característica de cantidad de colores

cantidad de imágenes con diagnósticos benignos, las gráficas muestran mayor densidad de puntos en esta sección. Por otro lado, se logra ver donde se encuentran ubicados los datos con mayor frecuencia, alrededor de la media de cada uno de los conjuntos, es decir, que si bien con esta gráfica no se puede apreciar un comportamiento notablemente diferente entre las dos clases si se puede observar cómo estos se encuentran distribuidos y concentrados alrededor de determinados valores, adicionalmente, concluir que si bien hay más datos de una clase que de otra, el comportamiento esperado según lo indicado en la regla ABCD se mantiene para las 7 características extraídas, como se puede ver en la figura 6-8.

Finalmente, las gráficas de densidad nos permiten observar el comportamiento, la distribución de los datos comparando dos conjuntos de características y la densidad de estos,

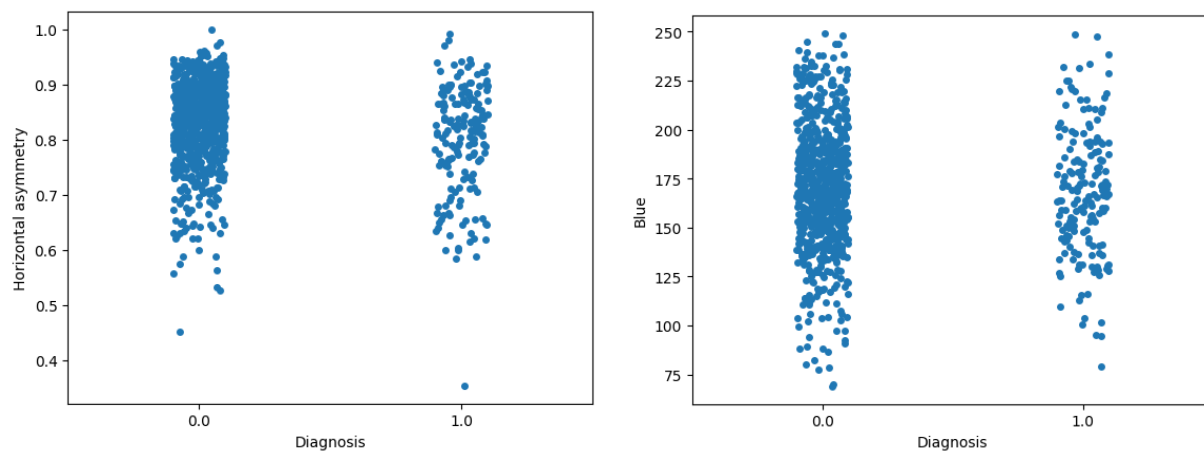


**Figura 6-7:** Diagrama de cajas del conjunto de entrenamiento para la característica de diámetro

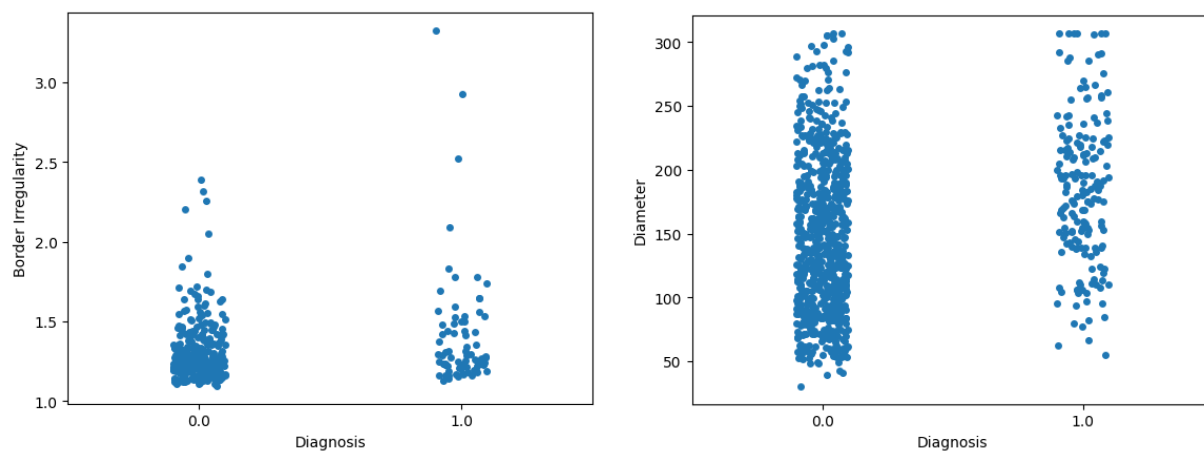
por ejemplo para la gráfica de asimetría horizontal con la cantidad de colores en canal rojo se observa que hay una diferencia en esta concentración de datos a pesar que su distribución es muy similar, ver figura 6-9-a, el conjunto de imágenes con etiqueta 1 se ubica en un valor mayor para la cantidad de rojos, pero menor para su asimetría horizontal, lo que está en sintonía con lo mencionado anteriormente con los diagramas de cajas. Otro ejemplo, es la gráfica de densidad de asimetría vertical y diámetro, figura 6-9-b, en la que las imágenes con melanoma se encuentran más concentradas en una posición superior, respecto al diámetro, en comparación con las imágenes benignas; más, sin embargo, no hay una diferencia significativa en la posición de esta densidad al mirar el eje correspondiente a la asimetría vertical.

## 6.2 Resultados del entrenamiento usando las características ABCD

Para la clasificación de melanoma usando las características ABCD, se hizo de dos formas; para la primera se usaron modelos clásicos de aprendizaje de máquina y para la segunda se validó con el uso de la librería *Lazy Predict*, para el primero de los casos se entrenaron 5 modelos diferentes con las imágenes de la base de datos (LR: Regresión logística, RF: *Random forest*, KNN: *K-nearest neighbor*, SVM: Máquinas de soporte vectorial, MLP: Perceptrón multicapa) y se obtuvieron métricas como la exactitud balanceada, sensibilidad, especificidad



(a) Gráfica de dispersión para la característica de asimetría horizontal (b) Gráfica de dispersión para la característica de colores - Canal azul



(c) Gráfica de dispersión para la característica de bordes irregulares (d) Gráfica de dispersión para la característica de diámetro

**Figura 6-8:** Gráficas de dispersión del conjunto de entrenamiento para cuatro de las siete características obtenidas

y el coeficiente Kappa, estos valores pueden observarse en las tablas **6-1** y **6-2**, de manera adicional se observó la matriz de confusión de cada uno de los modelos para visualizar el desempeño de los modelos durante la clasificación.

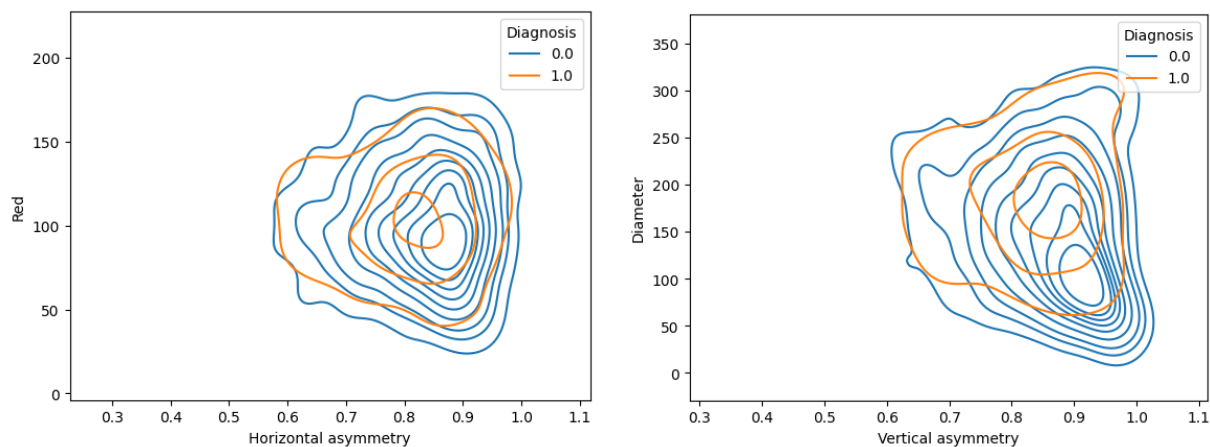
Para el modelo de regresión logística se obtuvo una buena clasificación de los datos con lesiones benignas, sin embargo, la gran mayoría de datos con melanoma las clasificó de manera incorrecta, como se puede apreciar en la figura **6-10** correspondiente a la matriz de confusión del modelo en prueba, de la cual se pueden obtener métricas como las mostradas en las tablas **6-1** y **6-2**, lo que en conjunto indica que este modelo identifica correctamente gran parte de los datos con lunares benignos, especificidad cercana a 1, sin embargo tiene una tasa alta de

| <b>Clasificador</b>  | <b>Balanced acc.</b> | <b>Sens.</b> | <b>Espec.</b> | <b>Kappa</b> |
|----------------------|----------------------|--------------|---------------|--------------|
| <b>LR</b>            | 0.5262               | 0.0635       | 0.9889        | 0.0795       |
| <b>KNN</b>           | 0.6475               | 0.7283       | 0.5667        | 0.1855       |
| <b>RF</b>            | 1.0                  | 1.0          | 1.0           | 1.0          |
| <b>SVM</b>           | 0.5                  | 1.0          | 0.0           | 0.0          |
| <b>MLP (64,256)</b>  | 0.5567               | 0.1271       | 0.9862        | 0.1644       |
| <b>MLP(256,512)</b>  | 0.5551               | 0.1213       | 0.9889        | 0.1616       |
| <b>MLP(512,1024)</b> | 0.5748               | 0.1676       | 0.9821        | 0.2104       |

**Tabla 6-1:** Métricas obtenidas en los 5 modelos clásicos en fase de entrenamiento

| <b>Clasificador</b>  | <b>Balanced acc</b> | <b>Sens.</b> | <b>Espec.</b> | <b>Kappa</b> |
|----------------------|---------------------|--------------|---------------|--------------|
| <b>LR</b>            | 0.5083              | 0.0266       | 0.9901        | 0.0259       |
| <b>KNN</b>           | 0.6314              | 0.7333       | 0.5296        | 0.0573       |
| <b>RF</b>            | 0.5534              | 0.12         | 0.9868        | 0.1551       |
| <b>SVM</b>           | 0.5                 | 1.0          | 0.0           | 0.0          |
| <b>MLP (64,256)</b>  | 0.5500              | 0.1066       | 0.9934        | 0.1485       |
| <b>MLP(256,512)</b>  | 0.5534              | 0.12         | 0.9868        | 0.1551       |
| <b>MLP(512,1024)</b> | 0.5517              | 0.12         | 0.9835        | 0.1492       |

**Tabla 6-2:** Métricas obtenidas en los 5 modelos clásicos en fase de prueba



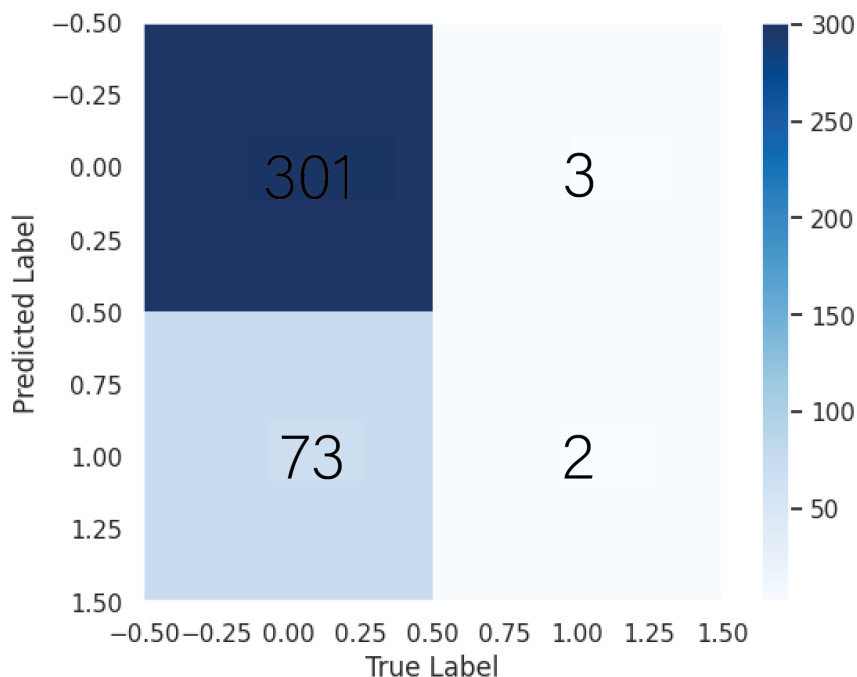
(a) Gráfica de densidad del conjunto de entrenamiento para las características de asimetría horizontal y colores en el canal rojo  
 (b) Gráfica de densidad del conjunto de entrenamiento para las características de asimetría vertical y diámetro

**Figura 6-9:** Gráfica de densidad para dos pares de características (Asimetría horizontal con colores en el canal rojo y asimetría vertical con diámetro)

falsos negativos, lo que quiere decir que clasifica como un lunar benigno algunos que tienen melanoma, esto se ve reflejado en el valor menor a 0.1 que tiene la sensibilidad. Todo lo anterior influye en que el valor del coeficiente *Kappa* indique un nivel de acuerdo ligero entre el modelo y el experto.

Por otro lado, el modelo de *K-nearest neighbor* presenta, a nivel general, un desempeño deficiente, tanto para clasificar melanoma, como para clasificar los datos con clasificación benignas, sin embargo, en comparación con el modelo de regresión logística este clasifica mejor la clase de melanoma, lo que se evidencia al comparar la matriz de confusión en la figura 6-11 con la figura 6-10 y observar el valor de la sensibilidad en entrenamiento, 0.7333. De manera detallada al observar la figura 6-11 se observa que para la clase 0, correspondiente a los datos con etiqueta de diagnóstico benigno, el modelo que emplea KNN clasifica el 56.6% de las imágenes de manera correcta, mientras que el restante 43.4% las clasifica como un falso positivo, mientras que, como se mencionó anteriormente, para la clase 1, lunares con melanoma, la diferencia entre verdaderos negativos y falsos negativos es mucho mayor (68.8% en comparación con 31.2%).

En tercera instancia, con el modelo de *Random Forest* se tuvo un excelente desempeño en entrenamiento, clasificó bien el 100% de los datos llegando a tener un coeficiente *Kappa* de 1, es decir un nivel de acuerdo casi perfecto con el experto. Sin embargo, con el conjunto de prueba el desempeño bajo bastante, clasificando correctamente 7 datos de melanoma como

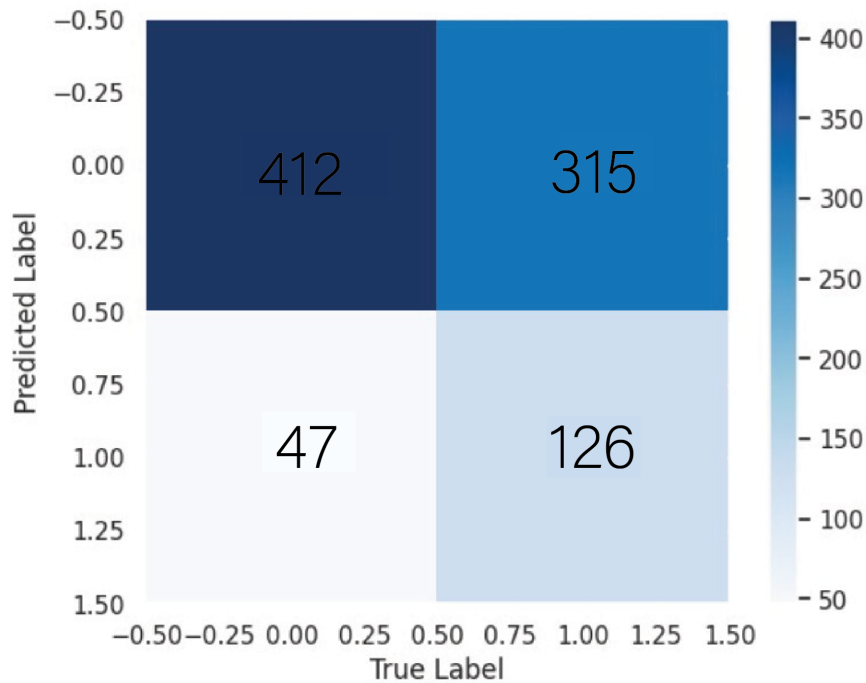


**Figura 6-10:** Matriz de confusión del modelo de regresión logística de las imágenes de prueba

se puede ver en la matriz de confusión de la figura **6-12**, llevando la sensibilidad a un valor de 0.12 y el coeficiente Kappa a 0.1551, que, si bien no es tan bajo como los dos modelos anteriores, si implica un nivel de acuerdo igual a estos, ligero. Esto se le puede atribuir a la clara tendencia que presento el modelo para clasificar correctamente unicamente la clase 0, mientras que para la clase 1 se observa una tasa alta de falsos negativos, es decir clasificar lunares con melanoma como lunares benignos.

Para el modelo de máquinas de soporte vectorial ocurrió el caso opuesto en la clasificación del modelo de regresión logística, todos los datos, tanto las de entrenamiento como las de prueba, fueron clasificadas como melanoma, ver figura **6-13**, el 100% de los datos clasificados como falsos positivos, lo que genera una especificidad y un coeficiente *Kappa* de 0, es decir, en este modelo no existe ningún nivel de acuerdo entre este mismo y el experto. Esto pudo haber ocurrido debido a que el conjunto de datos se encuentra en desbalance para la clase 1, lo que puede provocar que algoritmo al buscar el hiperplano lo haya más fácil al tener todos los datos clasificados en una sola clase.

Por último, dentro de la clasificación empleando modelos clásicos de aprendizaje profundo se usó perceptrón multicapa con tres configuraciones distintas de número de neuronas y de capas, pese a esto, los tres entrenamientos mostraron un comportamiento muy similar, con

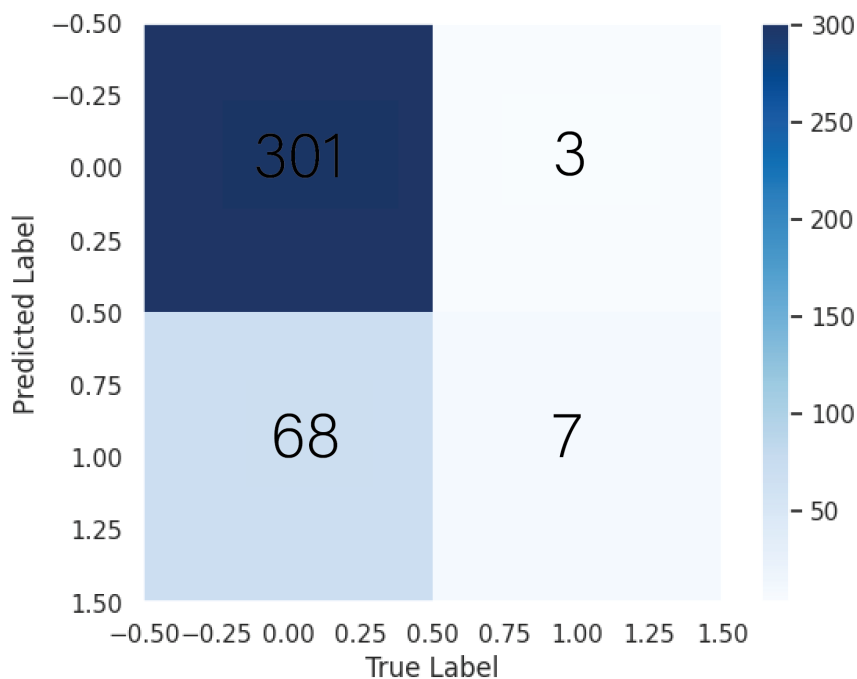


**Figura 6-11:** Matriz de confusión del modelo KNN de las imágenes de prueba

una exactitud balanceada de 0.56 aproximadamente en entrenamiento y prueba. Observando las matrices de confusión, figuras 6-14, es evidente que la clasificación tiende a los datos con lesiones benignas, ubicando muy pocas en la clase de melanoma y el resto como falsos negativos, esto implica que la sensibilidad para ninguna de las tres configuraciones sea mayor a 0.2 y que el coeficiente *Kappa* se encuentre entre 0.14 y 0.21 aproximadamente, un nivel de acuerdo entre ligero y justo.

Los resultados obtenidos empleando *lazy predict* permiten conocer cuales modelos presentan un mejor desempeño para clasificar el conjunto de datos sin realizar ajuste de parámetros [39], el uso de esta librería evaluó 26 modelos, de los cuales se seleccionaron los 4 mejores, ver figura 6-15 para emplear *GridSearch* y entrenar con los mejores parámetros.

Con estos 4 mejores modelos (*Nearest Centroid*, *Extra Tree Classifier*, *Label Propagation* y *Label Spreading*) se establecieron los parámetros mencionados en la tabla 5-1 y se entrenó con la mejor combinación de cada uno, ver tabla 5-2. A partir de este entrenamiento se obtuvieron las métricas mostradas en la 6-3, con estas se observa que para el modelo de *Nearest Centroid* la clasificación estuvo más balanceada para ambas clases, lo que se evidencia también de manera grafica con la matriz de confusión (figura 6-16), ambos valores de sensibilidad y especificidad son mayores a 0.60, sin embargo, el coeficiente *Kappa* indica que el nivel de acuerdo entre la clasificación del experto y la del modelo se mantiene en bajo.

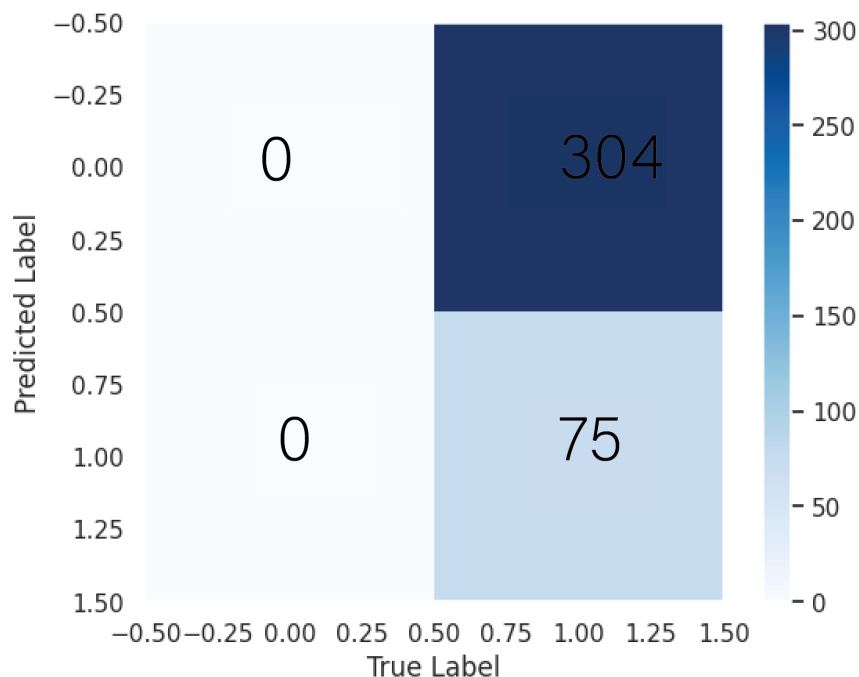


**Figura 6-12:** Matriz de confusión del modelo Random Forest de las imágenes de prueba

|                            | Balanced acc. | Sens.  | Espec. | Kappa  |
|----------------------------|---------------|--------|--------|--------|
| <b>NearestCentroid</b>     | 0.6109        | 0.6133 | 0.6085 | 0.1528 |
| <b>ExtraTreeClassifier</b> | 0.5134        | 0.04   | 0.9868 | 0.0407 |
| <b>LabelPropagation</b>    | 0.5186        | 0.08   | 0.9572 | 0.0518 |
| <b>LabelSpreading</b>      | 0.5186        | 0.0666 | 0.9671 | 0.0483 |

**Tabla 6-3:** Métricas de desempeño de los 4 mejores modelos de *Lazy Predict*

Esto en contraste con los otros tres modelos, que mostraron una clasificación de melanoma muy deficiente, prediciendo que gran parte de los datos correspondían a la etiqueta de lesión benigna, con valores para la métrica de sensibilidad inferiores a 0.1 (ver tabla 6-3 a pesar de clasificar de manera acertada la gran mayoría de datos con lunares benignos como se puede ver en la figura 6-17. Todo esto conlleva a que los valores del coeficiente *Kappa* sean muy cercanos a cero, lo que permite concluir que para estos modelos el nivel de acuerdo se encuentra muy cercano a pobre, de manera adicional, es importante resaltar que para estos tres modelos la cantidad de datos clasificados como benignos de manera incorrecta (falsos negativos) superan, en todos los casos, el 80%, lo que puede deberse a factores mencionados anteriormente como el desbalance de las clases, o la dificultad para encontrar diferencias significativas entre las características ABCD para las dos clases trabajadas.

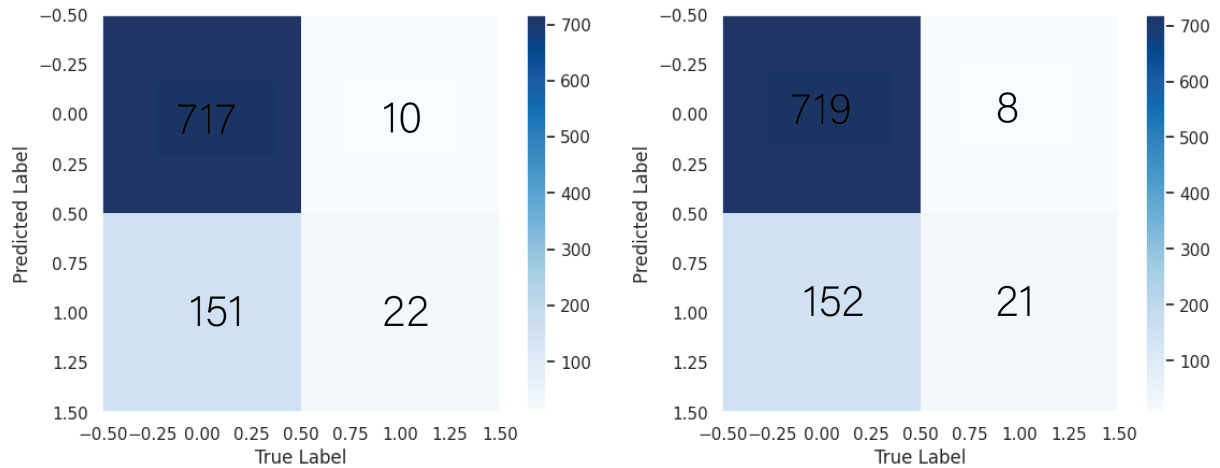


**Figura 6-13:** Matriz de confusión del modelo de máquinas de soporte vectorial de las imágenes de prueba

### 6.3 Resultados del entrenamiento usando las imágenes de la base de datos ISIC

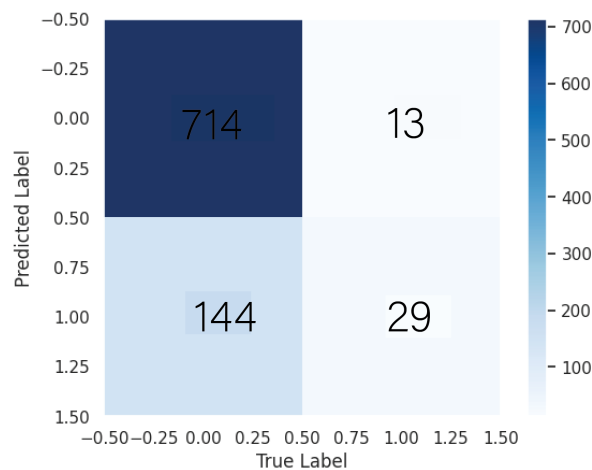
Para la clasificación de melanoma usando únicamente las imágenes, se realizaron las modificaciones mencionadas en el capítulo anterior, estas divididas en tres etapas y entrenando con dos conjuntos de datos, las imágenes del reto de 2016 y las del 2017 de la base de datos ISIC. Para la primera etapa donde se congelaron las capas, bajo ninguno de los experimentos se obtuvieron resultados positivos respecto a la clasificación de melanoma, las dos especificidades no tuvieron un valor mayor a 0.1, ver las tablas **6-4**, **6-6**, **6-5** y **6-7**, lo que conllevó a que el coeficiente *Kappa* no fuera mayor a 0.2. De manera adicional es importante resaltar que, para las imágenes de entrenamiento de 2017, ninguna de ellas fue clasificada como melanoma (ver matriz de confusión de la figura **6-18**), de aquí que la sensibilidad sea igual a 0 y el coeficiente *Kappa* tenga un valor menor a 0 (el menor valor obtenido durante el desarrollo de este proyecto), es decir, que no existe ningún nivel de acuerdo entre el experto y el modelo.

Los pesos de este modelo fueron usados en el segundo experimento, en el que se descongelaron las capas, y se realizó el entrenamiento. Para esta segunda etapa los resultados de los dos conjuntos de datos mostraron una mejoría en comparación con el modelo con capas congeladas, por ejemplo, se clasificaron correctamente todas las imágenes benignas del 2016



(a) Matriz de confusión del modelo perceptrón multicapa(64,256)

(b) Matriz de confusión del modelo perceptrón multicapa(256,512)



(c) Matriz de confusión del modelo perceptrón multicapa(512,1024)

**Figura 6-14:** Matrices de confusión de las tres configuraciones del modelo perceptrón multicapa para el conjunto de entrenamiento

como se puede ver en la figura 6-19, sin embargo, se sigue evidenciando dificultades para clasificar las imágenes de melanoma, por ejemplo, para el conjunto de prueba del año 2017 la sensibilidad tiene un valor de apenas 0.03 y un coeficiente de capa de 0.04, clasificando correctamente únicamente 4 imágenes de melanoma de 117 que son, ver figura 6-20.

Todo esto, aún después de realizado el aumento de los datos, es decir, para estos modelos el contar únicamente con las imágenes para la extracción de características y posteriormente su clasificación sigue siendo una tarea difícil de abordar, puesto que, como se mostró en la

| Model                  | Accuracy | Balanced Accuracy | ROC AUC | F1 Score | Time Taken |
|------------------------|----------|-------------------|---------|----------|------------|
| NearestCentroid        | 0.68     | 0.68              | 0.68    | 0.71     | 0.02       |
| ExtraTreeClassifier    | 0.75     | 0.62              | 0.62    | 0.75     | 0.01       |
| LabelPropagation       | 0.77     | 0.61              | 0.61    | 0.77     | 0.07       |
| LabelSpreading         | 0.77     | 0.61              | 0.61    | 0.77     | 0.09       |
| XGBClassifier          | 0.80     | 0.59              | 0.59    | 0.77     | 0.15       |
| DecisionTreeClassifier | 0.71     | 0.58              | 0.58    | 0.72     | 0.03       |
| GaussianNB             | 0.80     | 0.58              | 0.58    | 0.77     | 0.02       |
| LGBMClassifier         | 0.80     | 0.57              | 0.57    | 0.76     | 0.17       |

Figura 6-15: Tabla de los mejores modelos mostrada luego del uso de la librería *lazy predict*

|            | Balanced acc. | Sens.  | Espec. | Kappa  |
|------------|---------------|--------|--------|--------|
| Freeze     | 0.5414        | 0.0924 | 0.9903 | 0.1237 |
| Unfreeze 1 | 0.6300        | 0.2601 | 1.0    | 0.3622 |
| Unfreeze 2 | 0.8808        | 0.7630 | 0.9986 | 0.8352 |

Tabla 6-4: Métricas de desempeño para el conjunto de entrenamiento de la base de datos del año 2016

figura 5-5, las imágenes de ambas clases suelen presentar una similitud elevada entre ellas, lo que genera tener resultados como estos a pesar de haber usado técnicas para aumentar su precisión como el congelamiento de los pesos de ciertas capas.

Con estos resultados se vuelve a realizar el entrenamiento con las capas congeladas, pero con los pesos obtenidos del segundo experimento, y, al igual que sucedió con la etapa anterior, se evidencia una mejora significativa para la clasificación de las imágenes. De manera particular, el conjunto de entrenamiento del año 2017 aumento su exactitud balanceada de 0.5638 a 0.9434, logrando el mejor coeficiente *Kappa* hasta el momento, 0.87, un nivel de acuerdo casi perfecto (ver su matriz de confusión en la figura 6-21), observando los conjuntos de prueba se sigue evidenciando la mejora en las métricas de desempeño, pero persiste la dificultad para clasificar las imágenes con lunares malignos, por ejemplo, a pesar de que para las imágenes de 2016, la sensibilidad llego a 0.28, al observar la matriz de confusión, 6-22 resulta que fueron clasificadas como falsos negativos más de la mitad de las imágenes con melanoma.

|                   | <b>Balanced acc.</b> | <b>Sens.</b> | <b>Espec.</b> | <b>Kappa</b> |
|-------------------|----------------------|--------------|---------------|--------------|
| <b>Freeze</b>     | 0.5334               | 0.08         | 0.9868        | 0.0992       |
| <b>Unfreeze 1</b> | 0.5783               | 0.16         | 0.9967        | 0.2275       |
| <b>Unfreeze 2</b> | 0.6202               | 0.28         | 0.9605        | 0.3048       |

**Tabla 6-5:** Métricas de desempeño para el conjunto de prueba de la base de datos del año 2016

|                   | <b>Balanced acc.</b> | <b>Sens.</b> | <b>Espec.</b> | <b>Kappa</b> |
|-------------------|----------------------|--------------|---------------|--------------|
| <b>Freeze</b>     | 0.5250               | 0.0508       | 0.9993        | 0.0789       |
| <b>Unfreeze 1</b> | 0.5638               | 0.1283       | 0.9993        | 0.1919       |
| <b>Unfreeze 2</b> | 0.9434               | 0.9171       | 0.9698        | 0.8708       |

**Tabla 6-6:** Métricas de desempeño para el conjunto de entrenamiento de la base de datos del año 2017

|                   | <b>Balanced acc.</b> | <b>Sens.</b> | <b>Espec.</b> | <b>Kappa</b> |
|-------------------|----------------------|--------------|---------------|--------------|
| <b>Freeze</b>     | 0.4958               | 0.0          | 0.9917        | -0.0130      |
| <b>Unfreeze 1</b> | 0.5150               | 0.0341       | 0.9958        | 0.0469       |
| <b>Unfreeze 2</b> | 0.6746               | 0.4444       | 0.9047        | 0.3721       |

**Tabla 6-7:** Métricas de desempeño para el conjunto de prueba de la base de datos del año 2017

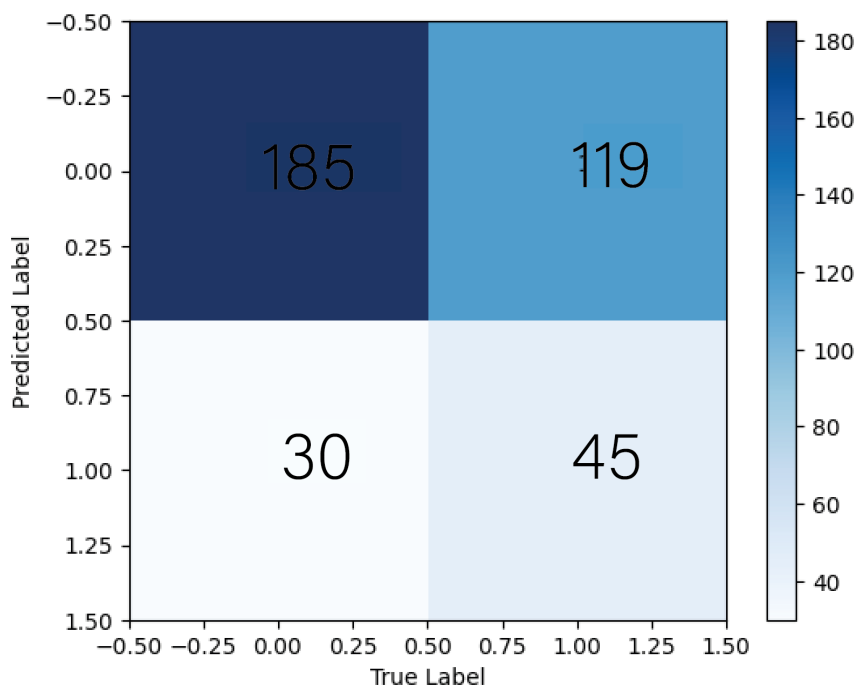
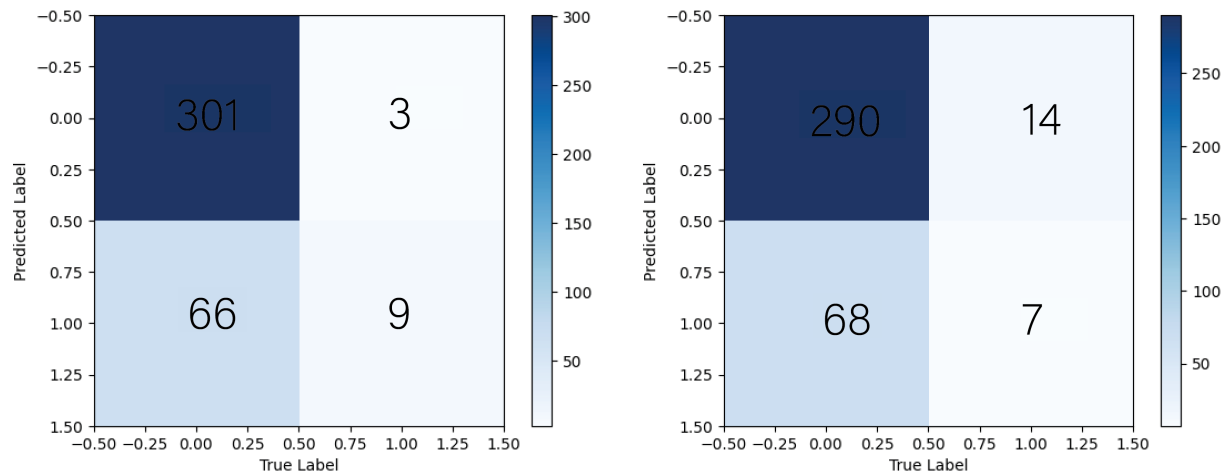


Figura 6-16: Matriz de confusión del modelo Nearest Centroid

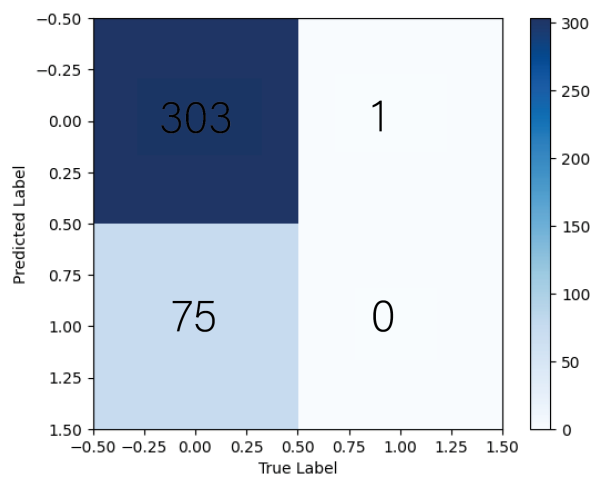
## 6.4 Resultados del entrenamiento usando las imágenes de la base de datos ISIC junto con sus características ABCD

Con base en los resultados obtenidos con el entrenamiento de los modelos con las imágenes y con las características, se realizó el entrenamiento de otros modelos que usen ambos conjuntos de datos, para ello se emplearon dos arquitecturas, *InceptionResNet* y *EfficientNet*, a cada una se le agregaron 3 capas adicionales explicadas en el capítulo anterior.

Para la primera arquitectura, con solo una capa adicional se tuvo una sensibilidad cerca de 1, específicamente 0.9768 como se puede ver en la tabla 6-8 para el conjunto de entrenamiento, es decir, realiza una buena clasificación de las imágenes de melanoma, pero no de las imágenes benignas, esto se ve sustentado con su matriz de confusión en la figura 6-23, donde se evidencia una tasa alta de falsos positivos contra los verdaderos positivos, a pesar de esto es importante resaltar que a comparación de los resultados obtenidos para los dos entrenamientos previos (usando únicamente el conjunto de datos de características ABCD y el entrenamiento con las imágenes) la clasificación de melanoma muestra un mejor balance general. Este comportamiento se mantiene para el conjunto de prueba, donde al ver la tabla 6-9, el índice *Kappa* toma un valor menor a cero, 0.0083, mostrando un nivel de acuerdo



(a) Matriz de confusión del modelo ExtraTreeClassifier (b) Matriz de confusión del modelo LabelPropagation

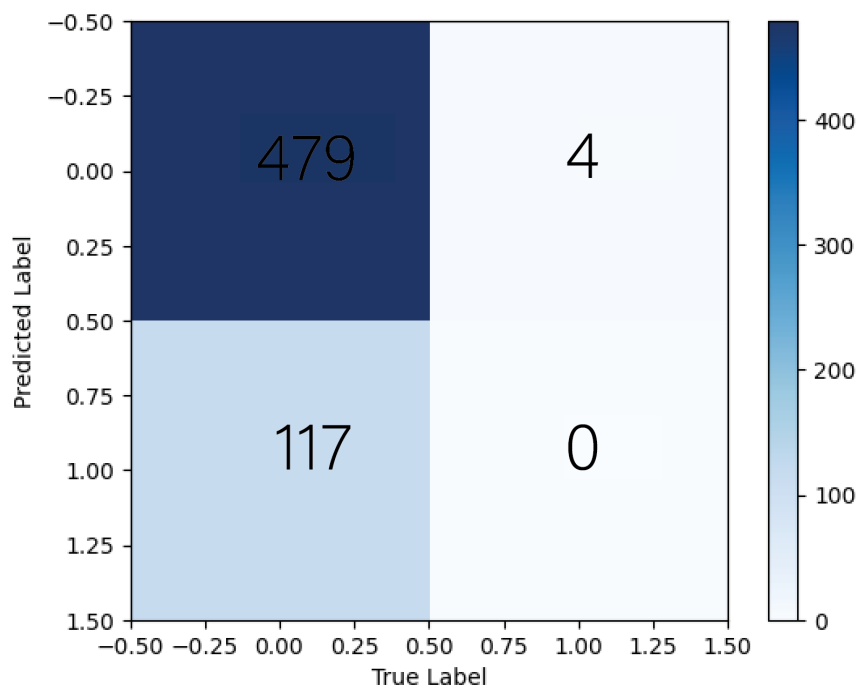


(c) Matriz de confusión del modelo LabelSpreading

**Figura 6-17:** Matrices de confusión de los otros tres mejores modelos de *Lazy Predict*

pobre con el experto, ya que no muestra una clasificación acertada para las imágenes con lunares benignos, esto puede deberse a que esta arquitectura suele presentar dificultades para generalizar ante nuevos conjuntos de datos, lo que la hace más susceptible a presentar sobreajustes.

Posterior a esto, se le adicionaron dos capas a la misma arquitectura, con esto se logró observar una mejora en la clasificación de imágenes con lesiones benignas, pero un retroceso para la otra clase, es decir, al analizar la matriz de confusión en la figura 6-24, observa un aumento de los falsos negativos al compararla con su conjunto de entrenamiento, figura 6-23. Esto es coherente con las métricas mostradas en las tablas 6-9 y 6-8, donde para



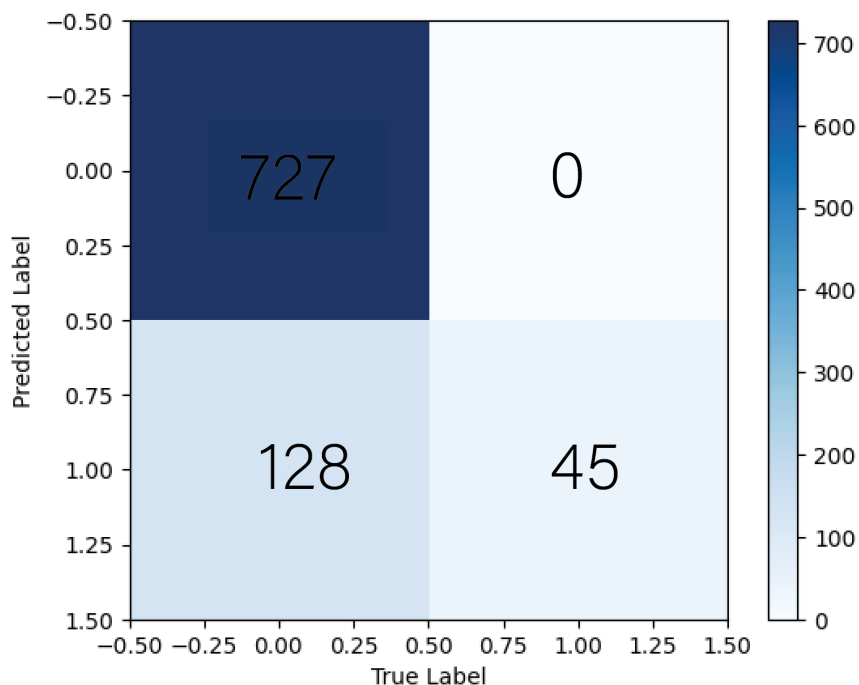
**Figura 6-18:** Matriz de confusión de la etapa de congelamiento de capas del conjunto de entrenamiento 2017

prueba la métrica de sensibilidad tiene un valor de 0.12 y el coeficiente *Kappa* permite una interpretación de acuerdo ligero con el experto. Este comportamiento fue muy similar con el modelo en el que se le agregaron 3 capas, es decir, clasifica bien la clase benigna pero presenta dificultades para la clasificación de la clase con melanoma, como se puede ver en la figura 6-25; sin embargo, el desempeño mejoró a comparación con el segundo experimento de esta arquitectura, en términos generales, para el conjunto de prueba se tuvo una exactitud balanceada de 0.5522.

Por otro lado, para la otra arquitectura, *EfficientNet*, se tuvieron mejores resultados en com-

| CNN             | Capas adicionales | Balanced acc. | Sens.  | Espec. | Kappa  |
|-----------------|-------------------|---------------|--------|--------|--------|
| InceptionResNet | 1                 | 0.5021        | 0.9768 | 0.0275 | 0.0017 |
|                 | 2                 | 0.5869        | 0.2427 | 0.9312 | 0.2117 |
|                 | 3                 | 0.6030        | 0.2543 | 0.9518 | 0.2600 |
| EfficientNet    | 1                 | 0.9467        | 0.9017 | 0.9917 | 0.9156 |
|                 | 2                 | 1.0           | 1.0    | 1.0    | 1.0    |
|                 | 3                 | 0.9913        | 0.9826 | 1.0    | 0.9891 |

**Tabla 6-8:** Métricas de desempeño para el conjunto de entrenamiento de la base de datos

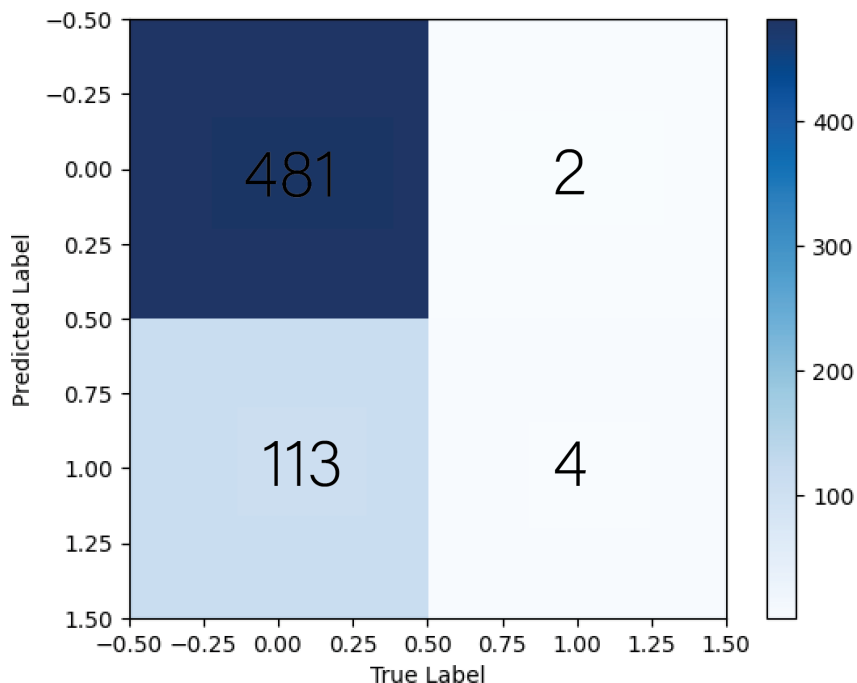


**Figura 6-19:** Matriz de confusión de la segunda etapa del conjunto de entrenamiento de 2016

paración con los obtenidos usando *InceptionResNet*, en entrenamiento los tres experimentos tuvieron una exactitud balanceada superior a 0.93, lo que indica que para este conjunto de datos la clasificación muestra un buen desempeño, con todos los valores de coeficiente *Kappa* mayores a 0.90. De manera particular, para el primer experimento, en prueba, la clasificación de melanoma no presenta un buen desempeño, como se puede ver en la matriz de confusión en la figura **6-26**, mostrando que clasifica aproximadamente el doble de los verdaderos negativos como falsos negativos, es decir imágenes con melanoma las predice dentro de la clase de lunares benignos.

| CNN                    | Capas adicionales | Balanced acc. | Sens.  | Espec. | Kappa   |
|------------------------|-------------------|---------------|--------|--------|---------|
| <b>InceptionResNet</b> | 1                 | 0.4897        | 0.9466 | 0.0328 | -0.0083 |
|                        | 2                 | 0.5303        | 0.12   | 0.9407 | 0.0800  |
|                        | 3                 | 0.5522        | 0.1866 | 0.9177 | 0.1274  |
| <b>EfficientNet</b>    | 1                 | 0.6256        | 0.3466 | 0.9046 | 0.2793  |
|                        | 2                 | 0.6987        | 0.4666 | 0.9309 | 0.4395  |
|                        | 3                 | 0.6353        | 0.32   | 0.9506 | 0.3303  |

**Tabla 6-9:** Métricas de desempeño para el conjunto de prueba de la base de datos



**Figura 6-20:** Matriz de confusión de la segunda etapa del conjunto prueba de 2017

Observando el segundo experimento, con las dos capas adicionales, en entrenamiento se tiene una clasificación y un nivel de acuerdo con el experto perfecta, clasifica correctamente todas las imágenes de la base de datos, ver figura 6-27, pero en prueba, este desempeño no se mantiene, la sensibilidad pasa de 1 a 0.46, pese a esto es importante resaltar, que el coeficiente Kappa de este experimento es el más alto dentro del conjunto de prueba, específicamente 0.4395, interpretando que existe un nivel de acuerdo moderado. Finalmente, con la tercera capa adicionada, el desempeño del modelo disminuye, por ejemplo, en prueba si bien la clasificación de las imágenes benignas se mantiene, para las imágenes con melanoma esta disminuye, la matriz de confusión de la figura 6-28, muestra que más de la mitad de esta clase son clasificadas de manera incorrecta, con una especificidad de 0.9506 y una sensibilidad de 0.32.

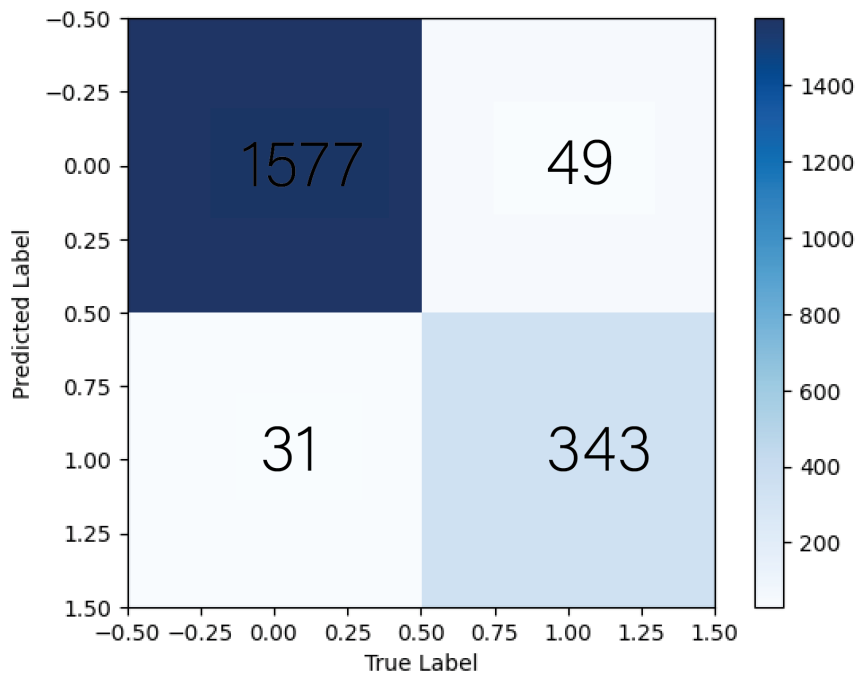


Figura 6-21: Matriz de confusión de la tercera etapa del conjunto de entrenamiento de 2017

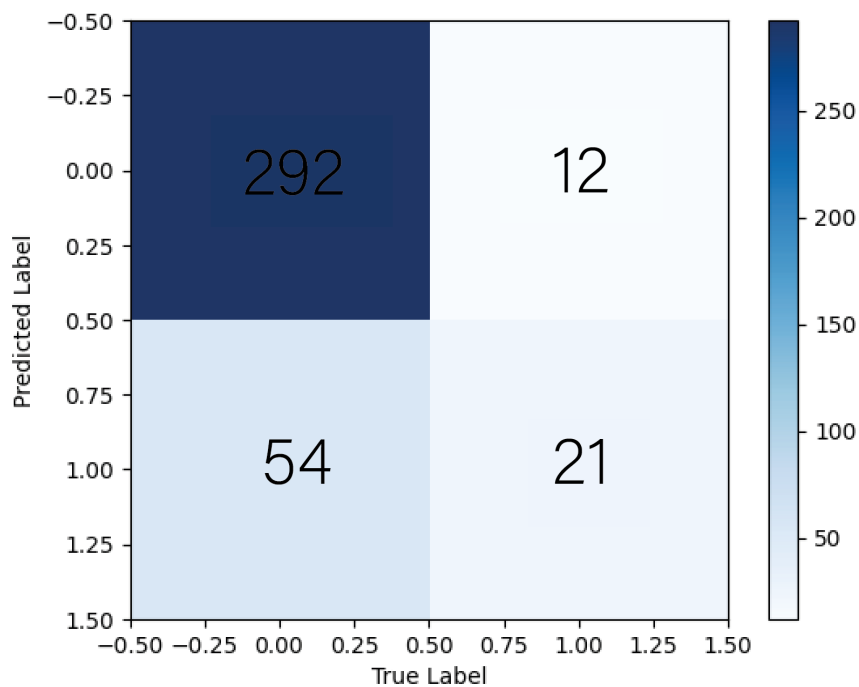
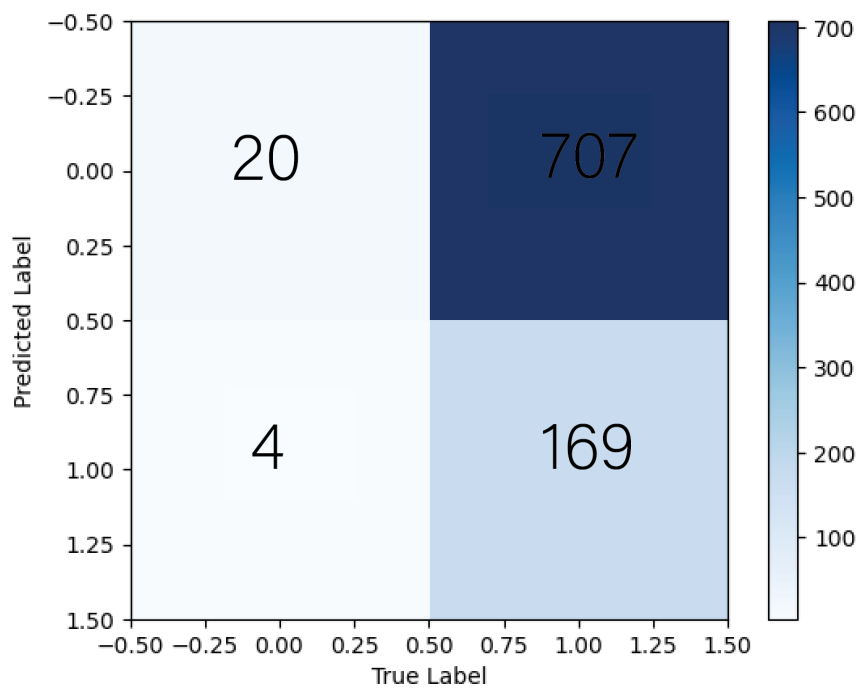
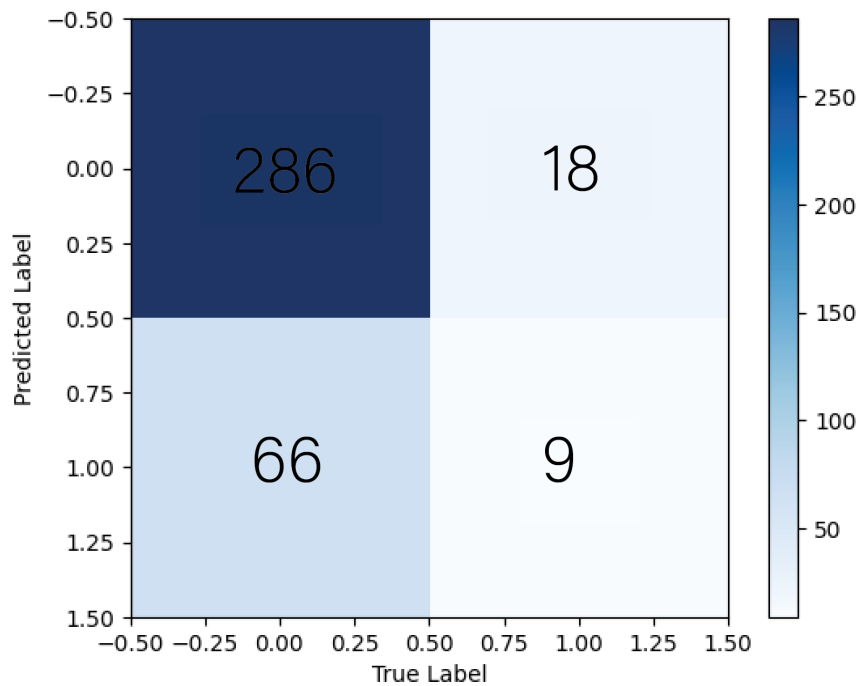


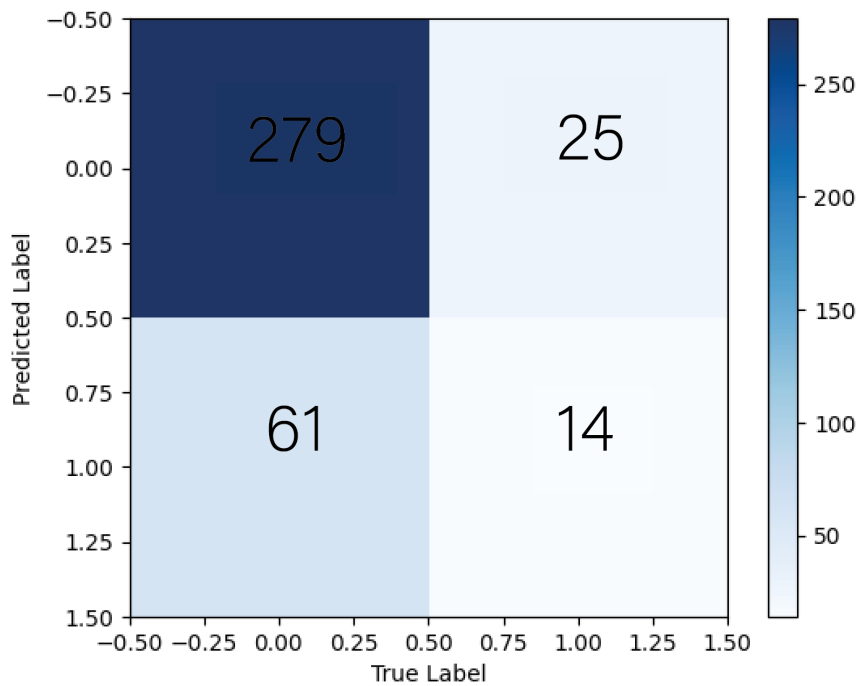
Figura 6-22: Matriz de confusión de la tercera etapa del conjunto prueba de 2016



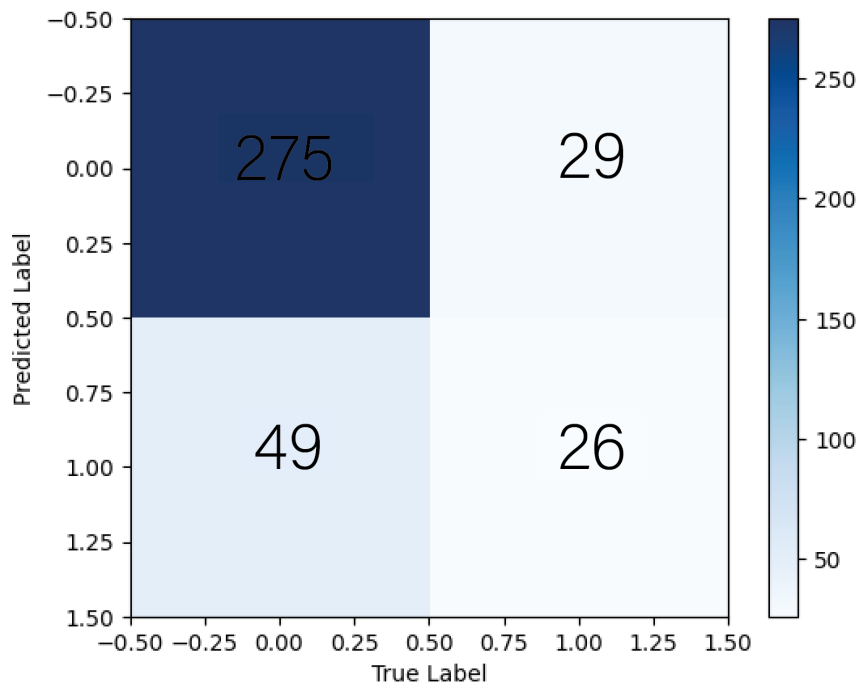
**Figura 6-23:** Matriz de confusión de la arquitectura InceptionResNet con una capa adicional del conjunto de entrenamiento



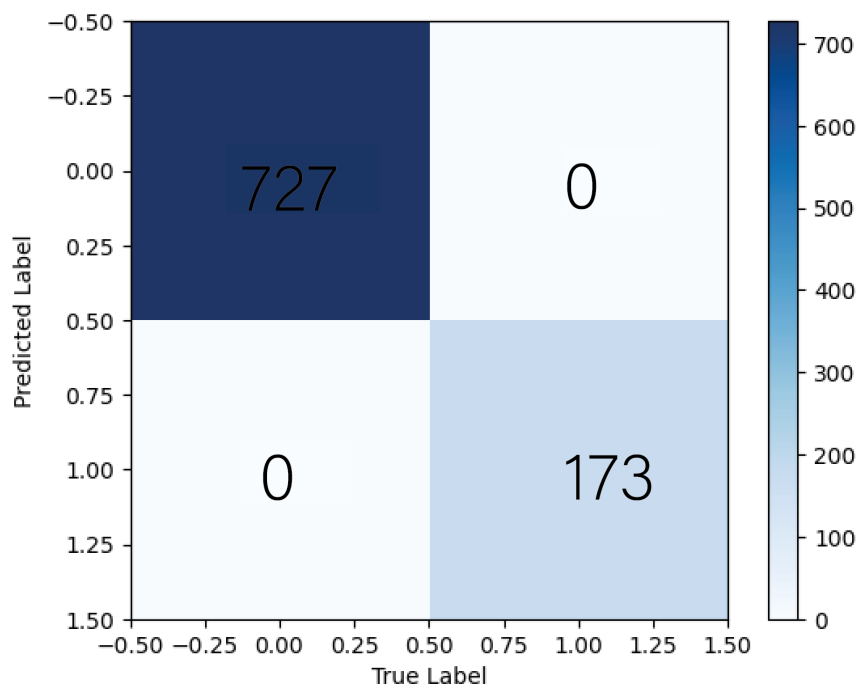
**Figura 6-24:** Matriz de confusión de la arquitectura InceptionResNet con dos capas adicionales del conjunto de prueba



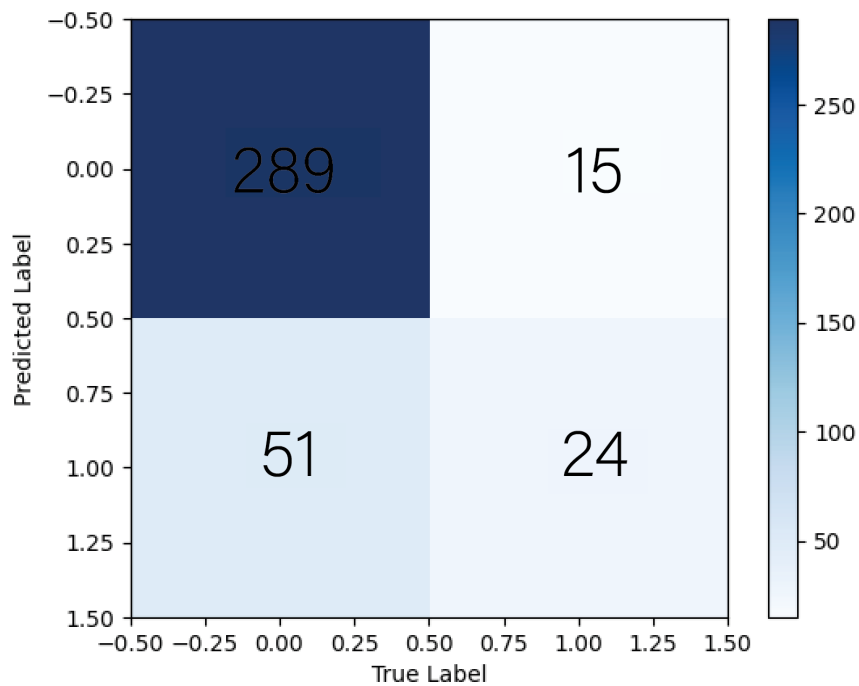
**Figura 6-25:** Matriz de confusión de la arquitectura InceptionResNet con tres capas adicionales del conjunto de prueba



**Figura 6-26:** Matriz de confusión de la arquitectura EfficientNet con una capa adicional del conjunto de prueba



**Figura 6-27:** Matriz de confusión de la arquitectura EfficientNet con dos capas adicionales del conjunto de entrenamiento



**Figura 6-28:** Matriz de confusión de la arquitectura EfficientNet con tres capas adicionales del conjunto de prueba

## 7 DISCUSIÓN

En este capítulo se hará la comparación de los resultados obtenidos entre ellos mismos y con los estudios similares descritos en la sección 2.1 de este documento. Con esto, se tiene que dentro de la revisión de literatura realizada se encontraron pocos estudios que combinen la regla ABCD junto con las imágenes, uno de estos es el realizado en 2019 por Moura et al. [34] quienes usaron esta regla como descriptor y el perceptrón multi capa junto con otras arquitecturas para realizar la clasificación, esto obtuvo que la mejor exactitud sea de 94.5% y el coeficiente Kappa de 89.2%; comparando esto con los resultados obtenidos con los modelos donde se emplean los dos conjuntos de datos (características ABCD e imágenes) se tiene que el modelo que usa la arquitectura *EfficientNet* junto con las dos capas adicionales de *GlobalAveragePooling2D* y *BatchNormalization* obtuvo mejores métricas de desempeño en entrenamiento, la cual tuvo una exactitud de 100% y un coeficiente *Kappa* igual 1. Lo anterior se le atribuye a que al tener dos conjuntos de características con las que entrenar (las siete de la regla ABCD y las extraídas automáticamente por el modelo) cuenta con más información para determinar patrones y características que estén más asociadas a una clase que a otra, sin embargo, este desempeño disminuyó en prueba, bajando su exactitud a un 70% aproximadamente y un coeficiente *Kappa* a 0.3303, lo que puede estar relacionado con la cantidad de imágenes que hay para la clase con melanoma.

Otro estudio que usa la regla ABCD junto con las imágenes para realizar la clasificación de melanoma es el desarrollado por Salma et al. [43], en este automatizan la extracción de características de la regla ABCD y luego hacen la clasificación usando redes neuronales convolucionales (CNN) combinadas con máquinas de soporte vectorial (SVM), con esto reportan que el mejor resultado tuvo una exactitud de 99.87% y una precisión de 98.77%. Al compararlo con los modelos ya mencionados, en términos generales, este estudio presenta un mejor desempeño en prueba, los resultados obtenidos de este proyecto muestran que la mejor exactitud balanceada alcanzada fue de 0.70; esta diferencia de desempeño puede atribuirse a que combinan dos técnicas para la clasificación, CNN y SVM, lo que permite que el modelo mejore su rendimiento y sea menos vulnerable a la entrada de información diferente. Pese a esto, al comparar estos modelos junto con los que usan solo las características de la regla ABCD o solo las imágenes, se observa un mejor desempeño general en el uso de los dos conjuntos, teniendo tres modelos, todos los entrenados con la arquitectura *EfficientNet*, en los que el entrenamiento alcanzó valores superiores 0.9 tanto en exactitud como en el coeficiente *Kappa*, lo que puede deberse a que la arquitectura empleada resulta ser más eficiente

y robusta que la *InceptionResNet*, por lo cual el adicionar capas, permite que este mejore su rendimiento para obtener los resultados mostrados en la tabla **6-8**.

Por otro lado, los modelos entrenados únicamente con imágenes tuvieron un mejor rendimiento al usar la base de datos ISIC del reto 2017, que cuenta con aproximadamente mil imágenes más que las del reto 2016, esto da paso a que el modelo al tener más datos de los cuales extraer características cuenta con más información para realizar la clasificación. El mejor de estos modelos alcanzó en prueba una sensibilidad de 0.4444 y una exactitud de 0.6746 pero en entrenamiento estas métricas fueron de 0.9171 y de 0.9434, respectivamente; a esto se le atribuye un sobreajuste del modelo posiblemente debido a su complejidad atribuida a la difícil tarea que resulta clasificar las imágenes pues ambas clases no presentan muchas diferencias en cuanto a características por extraer, como se puede ver en la figura **5-5**.

Al comparar estos modelos con resultados de estudios relacionados, se encuentra que el modelo propuesto por Codella et al. [10] en el que combinaron diferentes algoritmos basados en aprendizaje profundo con máquinas de soporte vectorial reporta una exactitud de 91.9% y una sensibilidad de 90.3%, que al comparar con el segundo modelo de capas descongeladas del conjunto de imágenes de 2017 reporta métricas menores, puestos tienen una exactitud de 94.34% y una sensibilidad de 91.71%, sin embargo, para el mismo modelo mencionado entrenado con el conjunto de imágenes de 2016 los resultados del estudio de Codella et al. presentaron mejores métricas, 88.08% de exactitud frente a la exactitud mencionada anteriormente, esto se debe a que la base de datos del reto de 2017 cuenta con más imágenes que la del reto de 2016, lo que permite que durante el entrenamiento las CNN tengan más información acerca de las características importantes para la clasificación binaria de imágenes y con esto detectar patrones relevantes.

El estudio realizado por Poorna et al. [40] se basa en comparar diferentes clasificadores clásicos usando imágenes de melanoma, incluyendo el puntaje dermatoscópicas total que emplea la regla ABCD sin técnicas de inteligencia artificial, este tuvo el menor desempeño mostrando una exactitud de 57.3%, observando aquellos que usaron diferentes técnicas de aprendizaje de máquina el menor valor lo tuvo el modelo de *K-nearest neighbor* con una exactitud de 66.9% mientras que las CNN tuvieron un valor de 90%, comparando esto con el resultado de los modelos entrenados con las imágenes de la base de datos gran parte de las métricas del segundo entrenamiento con capas descongeladas tuvieron valores superiores al reportado por el estudio para KNN, mientras que con la CNN únicamente el entrenamiento de la base de datos del 2017 superó el valor reportado de exactitud, específicamente 94.34%, este resultado se le atribuye al beneficio que trae el entrenar con ciertas capas congeladas y luego volver a entrenar con estas descongeladas, como lo es el mejorar el rendimiento, permitiendo que el modelo aprenda de las características específicas que diferencian ambas clases, o el reducir el posible riesgo de sobreajuste que puede generar el entrenar el mismo

modelo sin estas adaptaciones.

Acerca de estudios que realizaran clasificación automática usando exclusivamente las características obtenidas empleando la regla ABCD no se encontró mayor información, pero al comparar estos modelos con los otros entrenados durante el desarrollo de este proyecto, estos fueron los que peor desempeño mostraron, alcanzando como valor máximo para exactitud 0.6314. El desempeño de estos modelos no es excelente, debido a que, de las 7 características obtenidas, la única que presenta una diferencia significativa entre las clases es el diámetro, las otras no tienen una diferencia mayor a 1 unidad, por lo que el entrenar modelos para que encuentren patrones particulares basados en estas características de cada clase es una tarea complicada.

Adicional a esto al observar las tablas **7-1** , **7-2** donde se muestran los mejores resultados obtenidos para cada uno de los entrenamientos y pruebas realizadas. Resulta evidente que no hay un equilibrio general entre las métricas de sensibilidad y especificidad que resulta ser más notorio al observar los mejores resultados en prueba (tabla **7-2**). Lo anterior puede deberse a diversos motivos, por ejemplo, para el caso de los modelos entrenados únicamente con el conjunto de características extraídas , puede atribuirse a un sesgo de los datos, es decir, debido a que los datos entrenados en esta etapa del proyecto se encuentran en mayor proporción para la clase de lesiones benignas, esto genera que durante el entrenamiento el modelo sea más sensible a esta clase que a la otra, generando el desequilibrio que se aprecia en la tabla **7-2** para esta fase.

Por otro lado, para los otros entrenamientos este desequilibrio entre métricas se puede atribuir a la complejidad de los modelos empleados ya que como tal el conjunto de imágenes, bien sea de forma individual o, usada en conjunto con los datos extraídos para las características ABCD es un conjunto grande de datos que cuenta con características complejas de extraer de tal forma que se hallen diferencias significativas entre cada una de las clases. Adicionalmente, para la etapa donde se entrenó la red con los dos conjuntos de datos (imágenes y características ABCD) el modelo empleado resulta ser de una complejidad mayor que los otros modelos utilizados, por lo cual tiene una tendencia a presentar un sobreajuste durante el entrenamiento lo que perjudica el rendimiento del modelo en prueba, es decir cuando se le presentan datos nuevos, y por lo tanto el equilibrio entre sensibilidad y especificidad; esto se puede evidenciar al analizar ambas tablas, **7-1** , **7-2** y observar el equilibrio que existe entre estas dos métricas para el conjunto de entrenamiento sin embargo al observar estas mismas métricas para el conjunto de prueba se evidencia el desequilibrio mencionado atribuido a los argumentos dados anteriormente.

| <b>Clasificador</b>    | <b>Balanced acc</b> | <b>Sens</b> | <b>Espec</b> | <b>Kappa</b> |
|------------------------|---------------------|-------------|--------------|--------------|
| RF                     | 1.0                 | 1.0         | 1.0          | 1.0          |
| Unfreeze 2 (2016)      | 0.8808              | 0.7630      | 0.9986       | 0.8352       |
| Unfreeze 2 (2017)      | 0.9434              | 0.9171      | 0.9698       | 0.8708       |
| EfficientNet + 2 capas | 1.0                 | 1.0         | 1.0          | 1.0          |

**Tabla 7-1:** Tabla comparativa de los mejores modelos en cada fase para entrenamiento

| <b>Clasificador</b>    | <b>Balanced acc</b> | <b>Sens</b> | <b>Espec</b> | <b>Kappa</b> |
|------------------------|---------------------|-------------|--------------|--------------|
| RF                     | 0.5534              | 0.12        | 0.9868       | 0.1551       |
| Unfreeze 2 (2016)      | 0.6202              | 0.28        | 0.9605       | 0.3048       |
| Unfreeze 2 (2017)      | 0.6746              | 0.4444      | 0.9047       | 0.3721       |
| EfficientNet + 2 capas | 0.6987              | 0.4666      | 0.9309       | 0.4395       |

**Tabla 7-2:** Tabla comparativa de los mejores modelos en cada fase para prueba

## 8 RECOMENDACIONES Y TRABAJOS FUTUROS

A partir de los resultados obtenidos para los modelos de clasificación usando las características ABCD de la base de datos, se recomienda que para trabajos futuros se explore esta misma clasificación usando este conjunto de datos estandarizado con el fin de mejorar la precisión de estos modelos, pues deja todas las características en la misma escala, lo que beneficia el entrenamiento de los modelos de clasificación al no tener características en escalas notablemente distintas, como sucede con el caso de asimetría horizontal y vertical comparándola con el diámetro.

De manera adicional, para los modelos de clasificación usando únicamente las imágenes se pueden realizar exploraciones con arquitecturas diferentes a la empleada aca, o diferentes técnicas de clasificación que puedan combinarse para aumentar la precisión del modelo; también se sugiere implementar técnicas que permitan reducir el sesgo del modelo, que al no contar con un balance entre clases y que el total de datos puedan no ser representativos de la población se presenta una disminución de su precisión incluso después de emplear técnicas para aumento de datos.

Por otro lado, también se pueden explorar alternativas a la regla ABCD que toman en cuenta otras características, como la lista de 7 puntos que tiene en cuenta rasgos más específicos del lunar que el ojo no entrenado no puede distinguir con facilidad como, por ejemplo, observa si hay presencia de vasos sanguíneos irregulares en el lunar o si hay puntos de diferentes tamaños distribuidos irregularmente dentro y fuera del lunar [11]. El uso de algoritmos diferentes al empleado en este proyecto, para ser usado en conjunto o como alternativa a la regla ABCD puede mejorar el desempeño de los modelos entrenados para la clasificación de melanoma.

Finalmente, para el modelo de clasificación usando las imágenes y las características ABCD se puede realizar una exploración de parámetros de la arquitectura *EfficientNet* que fue la que presento mejores resultados, para tener un mejor rendimiento y desempeño del modelo. Adicionalmente, evaluar el uso de arquitecturas diseñadas para dispositivos móviles para poder emplearlas una siguiente fase del proyecto donde se plantee el desarrollo de un *software* de apoyo diagnóstico.

## 9 CONCLUSIONES

Se lograron obtener las características de asimetría, bordes, colores y diámetro de las imágenes de la base de datos, observando que si bien si existe una diferencia entre las dos clases evaluadas (etiquetas benignas y malignas) como era lo esperado, esta no es significativa para las cuatro primeras características, de manera específica el diámetro fue la única en la que se observó una diferencia notoria de los valores correspondientes a cada una de las clases.

Por otro lado, se logró hacer la adaptación de los modelos para extracción de características y clasificación de melanoma empleando datos diferentes de entrada: características ABCD, imágenes y los dos conjuntos de datos. Al analizar todos los modelos desarrollados usando todas las diferentes técnicas según los datos de entrada de estos, se encontró que aquellos donde se emplearon las imágenes en conjunto con las características fueron los que tuvieron mejores métricas de desempeño, particularmente los que usaron la arquitectura *Efficient-Net*. Esto se debe a que al usar dos conjuntos de datos se complementa la información, es decir, al agregar las características ABCD a las imágenes se le brinda al modelo una información adicional que junto con las extraídas automáticamente por este, permite que el modelo aproveche la información de los dos conjuntos de datos y de esta forma identifique representaciones más sólidas y distintivas de cada clase.

De manera adicional, se concluyó que el usar una mayor cantidad de datos para entrenar, así como realizar modificaciones a las capas de los modelos o a los hiper parámetros de entrenamiento afectan significativamente el desempeño del modelo y por lo tanto su habilidad para realizar clasificaciones correctas. Por este motivo se propuso como recomendaciones y trabajos futuros, realizar más exploraciones de este tipo con los modelos de mejor desempeño con el fin de obtener mejores métricas de prueba con los modelos.

Finalmente, la gran mayoría de los modelos desarrollados al validar su habilidad para realizar la clasificación, mostraron buenos resultados. Con excepción de los modelos donde se usó únicamente el conjunto de las características de la regla ABCD, donde la clasificación no tuvo el desempeño esperado es por esta razón que se motiva el uso de otros algoritmos diferentes, como la regla de los 7 puntos, la cual hace uso de otras características diferentes y suele ser más empleada dentro del campo médico que por personas sin experiencia, bien en conjunto o como alternativa a la regla ABCD, para determinar si esta mejora el rendimiento de los modelos adaptados para clasificación de lunares. Para los modelos en los que se usaron

las imágenes solas o en conjunto con las características de la regla ABCD para entrenar, se tuvo un buen balance general, es decir, rendimientos y desempeños buenos para clasificar las imágenes dermatoscópicas de lunares entre lesiones benignas o con melanoma, sin embargo, es importante evaluar otras adaptaciones de modelos para alcanzar bien sea el equilibrio entre sensibilidad y especificidad o evaluar los costos de estos falsos positivos o falsos negativos.

# Referencias

- [1] Álvaro Enrique Acosta et al. “Melanoma: patogénesis, clínica e histopatología”. In: *Revista de la Asociación Colombiana de Dermatología y Cirugía Dermatológica* 17.2 (2009). Number: 2, pp. 87–108. ISSN: 1657-0448. URL: <https://revista.asocolderma.org.co/index.php/asocolderma/article/view/156> (visited on 03/10/2023).
- [2] Abder-Rahman H. Ali, Jingpeng Li, and Guang Yang. “Automating the ABCD Rule for Melanoma Detection: A Survey”. In: *IEEE Access* 8 (2020). Conference Name: IEEE Access, pp. 83333–83346. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2991034.
- [3] *An Introduction to Statistical Learning*. An Introduction to Statistical Learning. URL: <https://www.statlearning.com> (visited on 11/05/2023).
- [4] “Artificial intelligence in healthcare: opportunities and risk for future”. In: *Gaceta Sanitaria* 35 (Jan. 1, 2021). Publisher: Elsevier Doyma, S67–S70. ISSN: 0213-9111. DOI: 10.1016/j.gaceta.2020.12.019. URL: <https://www.sciencedirect.com/science/article/pii/S0213911120302788> (visited on 11/02/2023).
- [5] Samia Benyahia, Boudjelal Meftah, and Olivier Lézoray. “Multi-features extraction based on deep learning for skin lesion classification”. In: *Tissue and Cell* 74 (Feb. 1, 2022). ISSN: 0040-8166. DOI: 10.1016/j.tice.2021.101701. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0040816621002172&site=eds-live> (visited on 04/21/2023).
- [6] Irene Boticonsejos. *La regla ABCDE para la prevención del melanoma ¿En qué consiste?* Boticonsejos. July 7, 2020. URL: <https://boticonsejos.es/la-regla-abcde-para-la-prevencion-del-melanoma/> (visited on 05/02/2023).
- [7] *cac\_2020\_11\_4\_consenso\_melanoma\_v6.pdf*. URL: [https://cuentadealtocosto.org/site/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https%3A%2F%2Fcuentadealtocosto.org%2Fsite%2Fwp-content%2Fuploads%2F2020%2F10%2Fcac\\_2020\\_11\\_4\\_consenso\\_melanoma\\_v6.pdf&download=false&print=true&openfile=false](https://cuentadealtocosto.org/site/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https%3A%2F%2Fcuentadealtocosto.org%2Fsite%2Fwp-content%2Fuploads%2F2020%2F10%2Fcac_2020_11_4_consenso_melanoma_v6.pdf&download=false&print=true&openfile=false) (visited on 03/10/2023).
- [8] *Cancer today*. URL: <http://gco.iarc.fr/today/home> (visited on 03/27/2023).
- [9] Hernán Carranza et al. “Genotipificación del melanoma en Colombia”. In: *Revista Colombiana de Hematología y Oncología* 2.3 (Oct. 1, 2013). Accepted: 2021-10-12T15:35:12Z Publisher: Asociación Colombiana de Hematología y Oncología, pp. 14–23. URL: <https://repositorio.fucsalud.edu.co/handle/001/1844> (visited on 10/31/2023).

- [10] Noel Codella et al. “Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images”. In: *Machine Learning in Medical Imaging*. Ed. by Luping Zhou et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 118–126. ISBN: 978-3-319-24888-2. DOI: 10.1007/978-3-319-24888-2\_15.
- [11] *Criterios dermatoscópicos para el diagnóstico del melanoma - Dialnet*. URL: <https://dialnet-unirioja-es.ez.urosario.edu.co/servlet/articulo?codigo=950555> (visited on 11/10/2023).
- [12] Charlotte Crowder. *Melanomas*. In: *Salem Press Encyclopedia of Health*. Journal Abbreviation: Salem Press Encyclopedia of Health. May 1, 2022. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=ers&AN=94462256&site=eds-live> (visited on 11/01/2023).
- [13] Thomas Davenport and Ravi Kalakota. “The potential for artificial intelligence in healthcare”. In: *Future Healthcare Journal* 6.2 (June 2019), pp. 94–98. ISSN: 2514-6645. DOI: 10.7861/futurehosp.6-2-94. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/> (visited on 11/02/2023).
- [14] *Definición de melanocito - Diccionario de cáncer del NCI - NCI*. Archive Location: nciglobal,ncicenterprise. Feb. 2, 2011. URL: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/melanocito> (visited on 10/28/2023).
- [15] *Definition of melanocyte - NCI Dictionary of Cancer Terms - NCI*. Archive Location: nciglobal,ncicenterprise. Feb. 2, 2011. URL: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/melanocyte> (visited on 10/11/2022).
- [16] Ahmet DEMİR and Feyza YILMAZ. “Inception-ResNet-v2 with Leakyrelu and Averagepooling for More Reliable and Accurate Classification of Chest X-ray Images”. In: *2020 Medical Technologies Congress (TIPTEKNO)*. 2020 Medical Technologies Congress (TIPTEKNO). ISSN: 2687-7783. Nov. 2020, pp. 1–4. DOI: 10.1109/TIPTEKNO50054.2020.9299232. URL: <https://ieeexplore.ieee.org/abstract/document/9299232> (visited on 11/06/2023).
- [17] *Día mundial del melanoma 2022*. Cuenta de Alto Costo. Section: Cáncer. May 19, 2022. URL: <https://cuentadealtocosto.org/site/cancer/dia-mundial-del-melanoma-2022/> (visited on 03/10/2023).
- [18] Eduardo Fierro et al. “Diagnóstico y tratamiento multidisciplinario de melanoma temprano y localmente avanzado. Consenso de expertos. Asociación Colombiana de Hemato-Oncología (ACHO).” In: *Revista Colombiana de Cancerología* 25.4 (Dec. 23, 2021). Number: 4, pp. 196–205. ISSN: 2346-0199. DOI: 10.35509/01239015.737. URL: <https://www.revistacancercol.org/index.php/cancer/article/view/737> (visited on 03/27/2023).

- [19] Alex M. Glazer et al. “Analysis of Trends in US Melanoma Incidence and Mortality”. In: *JAMA Dermatology* 153.2 (Feb. 1, 2017), pp. 225–226. ISSN: 2168-6068. DOI: 10.1001/jamadermatol.2016.4512. URL: <https://doi.org/10.1001/jamadermatol.2016.4512> (visited on 10/28/2023).
- [20] Jeremiah E. Golston et al. “Automatic detection of irregular borders in melanoma and other skin tumors”. In: *Computerized Medical Imaging and Graphics. Digital Imaging in Dermatology* 16.3 (May 1, 1992), pp. 199–203. ISSN: 0895-6111. DOI: 10.1016/0895-6111(92)90074-J. URL: <https://www.sciencedirect.com/science/article/pii/089561119290074J> (visited on 02/06/2024).
- [21] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: <http://www.deeplearningbook.org>.
- [22] Brent M. Gordon. *Artificial Intelligence: Approaches, Tools, and Applications*. Jan. 1, 2011. ISBN: 978-1-61324-019-9. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=440805&site=eds-live> (visited on 11/01/2023).
- [23] David Gutman et al. *Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)*. May 4, 2016. DOI: 10.48550/arXiv.1605.01397. arXiv: 1605.01397[cs]. URL: <http://arxiv.org/abs/1605.01397> (visited on 11/07/2023).
- [24] Russell C. Hardie et al. *Skin Lesion Segmentation and Classification for ISIC 2018 Using Traditional Classifiers with Hand-Crafted Features*. July 18, 2018. DOI: 10.48550/arXiv.1807.07001. arXiv: 1807.07001[cs, eess]. URL: <http://arxiv.org/abs/1807.07001> (visited on 04/21/2023).
- [25] International Agency for Research on Cancer Working Group on artificial ultraviolet (UV) light and skin cancer. “The association of use of sunbeds with cutaneous malignant melanoma and other skin cancers: A systematic review”. In: *International Journal of Cancer* 120.5 (Mar. 1, 2007), pp. 1116–1122. ISSN: 0020-7136. DOI: 10.1002/ijc.22453.
- [26] *ISIC — International Skin Imaging Collaboration*. ISIC. URL: <https://www.isic-archive.com/> (visited on 10/24/2023).
- [27] *ISIC Archive*. URL: <https://challenge.isic-archive.com/> (visited on 10/19/2022).
- [28] Mohamed A. Kassem, Khalid M. Hosny, and Mohamed M. Fouad. “Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning”. In: *IEEE Access* 8 (2020). Conference Name: IEEE Access, pp. 114822–114832. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.3003890.

- [29] Miroslav Kubat. *An Introduction to Machine Learning*. Cham: Springer International Publishing, 2017. ISBN: 978-3-319-63913-0. DOI: 10.1007/978-3-319-63913-0. URL: <http://link.springer.com/10.1007/978-3-319-63913-0> (visited on 11/05/2023).
- [30] Yann LeCun, Y. Bengio, and Geoffrey Hinton. “Deep Learning”. In: *Nature* 521 (May 28, 2015), pp. 436–44. DOI: 10.1038/nature14539.
- [31] Masoud Mahdianpari et al. “Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery”. In: *Remote Sensing* 10 (July 16, 2018), p. 1119. DOI: 10.3390/rs10071119.
- [32] Yasuhiro Matsumura and Honnavara N. Ananthaswamy. “Toxic effects of ultraviolet radiation on the skin”. In: *Toxicology and Applied Pharmacology* 195.3 (Mar. 15, 2004), pp. 298–308. ISSN: 0041-008X. DOI: 10.1016/j.taap.2003.08.019.
- [33] Iván Montenegro Trujillo et al. “Macrotendencias hacia el 2030: El mundo y América Latina”. In: <https://minciencias.gov.co/portafolio/unidad-politica/lineas-trabajo/documentos-politica-ctei> (Nov. 2018). Accepted: 2022-10-18T13:46:17Z Publisher: Colciencias. URL: <https://repositorio.minciencias.gov.co/handle/20.500.14143/44997> (visited on 11/02/2023).
- [34] Nayara Moura et al. “ABCD rule and pre-trained CNNs for melanoma diagnosis”. In: *Multimedia Tools and Applications* 78.6 (Mar. 1, 2019), pp. 6869–6888. ISSN: 1573-7721. DOI: 10.1007/s11042-018-6404-8. URL: <https://doi.org/10.1007/s11042-018-6404-8> (visited on 03/10/2023).
- [35] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. Google-Books-ID: RC43AgAAQBAJ. MIT Press, Sept. 7, 2012. 1102 pp. ISBN: 978-0-262-30432-0.
- [36] N. Gupta and R. Mangla. *Artificial Intelligence Basics : A Self-Teaching Introduction*. Jan. 1, 2020. ISBN: 978-1-68392-516-3. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=2387284&site=eds-live> (visited on 11/01/2023).
- [37] Reme Navarro. *¿Cómo distinguir un lunar normal de un melanoma?* Consejos de Farmacia Online Atida. June 20, 2019. URL: <https://www.atida.com/es-es/blog/2019/06/como-distinguir-un-lunar-normal-de-un-melanoma/> (visited on 06/20/2023).
- [38] Chan-Woo Park et al. “Artificial Intelligence in Health Care: Current Applications and Issues”. In: *Journal of Korean Medical Science* 35.42 (Oct. 27, 2020). Publisher: The Korean Academy of Medical Sciences. DOI: 10.3346/jkms.2020.35.e379. URL: <https://synapse.koreamed.org/articles/1146221?viewtype=pubreader> (visited on 11/02/2023).

- [39] Ronil Patil. *Lazy Predict - Best Suitable Model for You*. Analytics Vidhya. May 6, 2021. URL: <https://www.analyticsvidhya.com/blog/2021/05/lazy-predict-best-suitable-model-for-you/> (visited on 11/06/2023).
- [40] S. S. Poorna et al. “Computer Vision Aided Study for Melanoma Detection: A Deep Learning Versus Conventional Supervised Learning Approach”. In: *Advanced Computing and Intelligent Engineering*. Ed. by Bibudhendu Pati et al. Advances in Intelligent Systems and Computing. Singapore: Springer, 2020, pp. 75–83. ISBN: 9789811510816. DOI: 10.1007/978-981-15-1081-6\_7.
- [41] Adam I. Riker, ed. *Melanoma: A Modern Multidisciplinary Approach*. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-78310-9. DOI: 10.1007/978-3-319-78310-9. URL: <http://link.springer.com/10.1007/978-3-319-78310-9> (visited on 10/29/2023).
- [42] Yasmín Rodríguez Acosta et al. “Cresta neural, cuarta hoja, germinativa embrionaria”. In: *Revista Médica Electrónica* 42.4 (Aug. 2020). Publisher: 1997, Centro Provincial de Información de Ciencias Médicas de Matanzas–FCMM, pp. 2049–2065. ISSN: 1684-1824. URL: [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S1684-18242020000402049&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1684-18242020000402049&lng=es&nrm=iso&tlng=es) (visited on 10/31/2023).
- [43] Wessam Salma and Ahmed S. Eltrass. “Automated deep learning approach for classification of malignant melanoma and benign skin lesions”. In: *Multimedia Tools and Applications* 81.22 (Sept. 1, 2022), pp. 32643–32660. ISSN: 1573-7721. DOI: 10.1007/s11042-022-13081-x. URL: <https://doi.org/10.1007/s11042-022-13081-x> (visited on 01/31/2023).
- [44] Fernando Ferreira Lima dos Santos et al. “Open source iterative bayesian classifier algorithm for quality assessment of processed coffee beans”. In: *Nativa* 8.1 (Feb. 5, 2020). Number: 1, pp. 118–123. ISSN: 2318-7670. DOI: 10.31413/nativa.v8i1.8074. URL: <https://periodicoscientificos.ufmt.br/ojs/index.php/nativa/article/view/8074> (visited on 11/08/2023).
- [45] William H. Sharfman. *Melanoma*. Vol. 00003. Jan. 1, 2012. ISBN: 978-1-936287-79-6. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=529438&site=eds-live> (visited on 03/09/2023).
- [46] *Skin*. Servier Medical Art. URL: [https://smart.servier.com/smart\\_image/skin/](https://smart.servier.com/smart_image/skin/) (visited on 10/31/2023).
- [47] Kousai Smeda. *Understand the architecture of CNN*. Medium. Nov. 19, 2019. URL: <https://towardsdatascience.com/understand-the-architecture-of-cnn-90a25e244c7> (visited on 11/07/2023).
- [48] MD Stephan P. Ariyan MBA and MD Harriet Kluger. *The Melanoma Handbook*. Jan. 1, 2017. ISBN: 978-1-62070-118-8. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=1491861&site=eds-live> (visited on 03/10/2023).

- [49] Mingxing Tan and Quoc V. Le. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. version: 5. Sept. 11, 2020. DOI: 10.48550/arXiv.1905.11946. arXiv: 1905.11946[cs,stat]. URL: <http://arxiv.org/abs/1905.11946> (visited on 11/06/2023).
- [50] Soheil Tavakolpour, Maryam Daneshpazhooch, and HamidReza Mahmoudi. *Skin Cancer: Genetics, Immunology, Treatments, and Psychological Care*. Cancer Genetics and Psychotherapy. Pages: 934. Cham: Springer International Publishing, Jan. 1, 2017. 851 pp. ISBN: 978-3-319-64548-3. DOI: 10.1007/978-3-319-64550-6\_18.
- [51] James Thorn. *Logistic Regression Explained*. Medium. Sept. 26, 2021. URL: <https://towardsdatascience.com/logistic-regression-explained-9ee73cede081> (visited on 11/05/2023).
- [52] Esther de Vries. “Melanomas in Colombia; a different reality”. In: *Revista Colombiana de Cancerología* 17.3 (Sept. 1, 2013). Publisher: Elsevier, pp. 91–92. ISSN: 0123-9015. DOI: 10.1016/S0123-9015(13)70010-6. URL: <https://www.elsevier.es/es-revista-revista-colombiana-cancerologia-361-articulo-melanomas-in-colombia-different-reality-S0123901513700106> (visited on 10/29/2023).
- [53] Renjie Xu et al. “A Forest Fire Detection System Based on Ensemble Learning”. In: *Forests* 12 (Feb. 13, 2021), p. 217. DOI: 10.3390/f12020217.
- [54] Abdurrahim Yilmaz et al. *Benchmarking of Lightweight Deep Learning Architectures for Skin Cancer Classification using ISIC 2017 Dataset*. Oct. 23, 2021. DOI: 10.48550/arXiv.2110.12270. arXiv: 2110.12270[cs,eess]. URL: <http://arxiv.org/abs/2110.12270> (visited on 04/21/2023).