

**Universidad Autónoma de Nuevo León**

**FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS**

**MAESTRÍA EN CIENCIA DE DATOS**

*Detección de Melanoma en imágenes dermatoscópicas con diferentes arquitecturas*

Autor:

Lic. Leobardo García Reyes

---

Supervisado por:

Dr. Mayra Cristina Berrones Reyes

24 de julio del 2022

# Detección de Melanoma en imágenes dermatoscópicas con diferentes arquitecturas

*Resumen— El melanoma es una forma de cáncer de piel mortal, en donde, la exposición constante al sol (UV) es el principal factor de riesgo para la aparición de esta. Si no se trata a tiempo el melanoma, puede extenderse rápidamente a otros órganos. Afortunadamente, los síntomas del melanoma se vuelven visibles para las víctimas, lo que crea una oportunidad para detectarlo en una etapa temprana.*

*Se han propuesto diferentes métodos de CNN (Red Neuronal Convolutacional) para la detección de melanoma en imágenes dermatoscópicas, ya sea construidas propiamente o con Transfer Learning. También, usando técnicas de aumento de datos para contraste y reducir el ruido para que de esta forma pueda extraer la información más relevante*

*En este trabajo busca utilizar diferentes arquitecturas de Transfer Learning como ResNet50 V2, Xception, VGG16, MobileNet V2 e Inception V3 para compararlos entre sí y observar cuál de estos obtiene un mejor resultado.*

*Palabras clave—Melanoma, Transfer Learning, CNN*

## 1. Introducción

El melanoma es una neoplasia maligna de los melanocitos, células productoras de pigmento de origen neuroectodérmico que se pueden encontrar en todo el cuerpo (incluso en la piel, el iris y el recto). La exposición al sol (UV) es el principal factor de riesgo para el melanoma cutáneo y conduce a una firma genética que es característica del melanoma [4, 16].

Si no se trata a tiempo el melanoma, puede extenderse rápidamente a otros órganos. Afortunadamente, los síntomas del melanoma se vuelven visibles para las víctimas, lo que crea una oportunidad para detectarlo en una etapa temprana. Debido a que la gente sabe menos sobre los síntomas específicos y debido a la escasez de médicos expertos, la automatización de la detección del melanoma se ha convertido en un tema importante para la salud pública [14, 13].

La alta morbilidad y el considerable costo de atención médica asociados con el tipo de lesión maligna ha motivado a los investigadores a desarrollar algoritmos más precisos y flexibles para la detección temprana del melanoma. Las tasas de melanoma han ido en aumento durante los últimos 30 años. Sin embargo, el melanoma es altamente curable si se diagnostica en su etapa inicial [7].

El empleo de técnicas no invasivas ayuda en la recopilación de la información sobre la estructura y la forma del melanoma. El instrumento dermoscopio captura imágenes dermatoscópicas generalmente tienen una mejor invariancia a la iluminación y el contraste, pero las diferentes condiciones y parámetros de las imágenes, como distancias, iluminación no ideal, resolución deficiente y también limitaciones en la calidad de la imagen, como ruidos, sombras, artefactos, pelos y reflejos, disminuyen la calidad de la imagen, lo que podría comprometer la solidez del análisis de lesiones cutáneas [7].

Con las deficiencias que se pueden encontrar en las imágenes dermatoscópicas una de las soluciones es diseñar un algoritmo de CNN (Red Neuronal Convolutacional) propia o a través de Transfer Learning que pueda corregir el color, mejorar el contraste y reducir el ruido para que de esta forma pueda extraer la información más relevante siendo capaz de identificar y clasificar imágenes dermatoscópicas entre Melanoma y No Melanoma.

## 2. Estado del arte

Los avances en CNN ha demostrado resultados muy prometedores en tareas de reconocimiento de objetos y se ha convertido en un dominio de investigación exigente para la clasificación de imágenes médicas [7]. En el artículo [13], utiliza la técnica de aumento de datos para tener las dos clases balanceadas, ya que originalmente, se tenían más imágenes de No Melanoma. Se estudia el rendimiento de diferentes arquitecturas de CNN de Transfer Learning para la detección de melanoma y presenta una visión comparativa de esos estudios. Se probaron 4 modelos de Transfer Learning y 3 CNN de segmentación: **Transfer Learning:** ResNet50, InceptionV3, Xception y VGG16. **CNN segmentación:** SegNet, BCDU-Net y U-Net.

También, en el artículo [12], se propone una CNN liviana, el cual consiste en que después de la capa de imagen de entrada en la capa superior, hay tres capas de convolución 2D y están separadas por dos capas de grupo máximo en el medio y, finalmente, las capas completamente conectadas y la capa de máximo suave están presentes para la clasificación. A las imágenes originales se aumentan haciendo transformaciones de giro horizontal y vertical e introduciendo 4 tipos de ruidos, Gaussiano, Poisson, Salt-and-Pepper y moteado. Otras dos formas de aumento realizadas son el incremento y la disminución del brillo en un valor de 50 en la imagen original.

El trabajo [1] propone una CNN que tiene 17 capas y las capas se distribuyen en 5 bloques. Cada bloque tiene las siguientes capas: convolucional, ReLU, pooling y dropout. Excepto por el último bloque que contiene la capa totalmente conectada y la capa softmax para la clasificación. En general, la red tiene 5 capas convolucionales, 4 capas ReLU, 3 capas de agrupación máxima y 4 capas de abandono. Esta CNN es modificada de LightNet, donde estas imágenes se clasifican sin aplicar segmentación de lesiones ni pre-procesamiento complejo solo re-dimensionando las imágenes a un tamaño de  $256 \times 256$ .

El aporte del artículo [7], fue examinar el desempeño de 5 CNN de última generación, AlexNet, VGG-Net16, VGGNet19, ResNet50 y Xception con el objetivo de comprobar la efectividad de diversas técnicas de pre-procesamiento y aumento de datos. Se agregaron dos capas completamente conectadas para combinar los mapas de características de varias capas intermedias de CNN. Para ambas capas, se emplea el método de dropout que evita el problema de sobreajuste.

En el estudio de [14], se analizó cómo el rendimiento de la clasificación se vio afectado por la degradación y los defectos en las imágenes de lesiones. Consideraron distorsiones de contraste, ruido, desenfoco y objetos externos adicionales como cabello y regla. Usaron 4 arquitecturas CNN diferentes que se usaban comúnmente para el problema de clasificación del melanoma: Xception e Inception, ResNet V2, DenseNet121 y DenseNet201.

### 3. Materiales y Metodología

#### 3.1. Materiales

El conjunto de datos que se seleccionó para el entrenamiento del algoritmo, se obtuvo a través del sitio web "Kaggle" [11]. Esta misma página hace referencia a otro sitio web llamado "HARVARD Dataverse" [21] que lleva por título "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", el cual es el repositorio oficial de los datos. Se recolectaron imágenes dermatoscópicas provenientes de diferentes poblaciones que fueron almacenadas en diferentes modalidades.

El conjunto de datos final contiene 10,015 imágenes dermatoscópicas divididas en categorías importantes en el ámbito de las lesiones pigmentadas con su perteneciente cantidad de imágenes, en donde, el diagnóstico real fue validado por histopatología (también conocida como fuente de la verdad) en más del 50 % de los casos [10]:

- 327 imágenes de queratosis solares, carcinoma intraepitelial, enfermedad de Bowen (akiec).
- 514 imágenes de carcinoma de células basales (bcc).
- 1,099 imágenes de lesiones tipo queratosis benignas (léntigos solares/queratosis seborreicas y liquen plano tipo queratosis, bkl).
- 115 imágenes de dermatofibroma (df).
- 6,705 imágenes de nevos melanocíticos (nv).
- 1,113 imágenes de melanoma (mel).
- 142 imágenes de lesiones vasculares (angiomas, angioqueratomas, granulomas piógenos y hemorragia, vasc).

Se puede observar (Fig. 1), que el conjunto de datos está desequilibrado, ya que el número de imágenes en cada clase varía de 115 (dermatofibroma) a 6,705 (nevos melanocíticos). En la página [10], se decidió realizar una clasificación entre Melanoma y No Melanoma, pero el resultado de esta nueva clasificación sigue siendo un conjunto de datos desequilibrado con:

- 1,113 imágenes de melanoma.
- 8,902 imágenes de no melanoma.

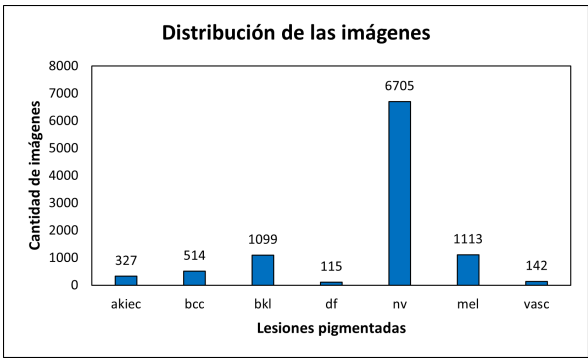


Figura 1: Distribución de las imágenes del conjunto de datos. Fuente: Imagen de elaboración propia.

Entonces, en la misma página se aplicó un aumento de datos para la clase de Melanoma para que sea la misma cantidad de la clase No Melanoma. En este aumento de datos, se utilizaron modificaciones de imagen como zoom, ángulo, desplazamiento, giros horizontales y verticales, etc., y crea imágenes sintéticas [10].

Las imágenes previamente desequilibradas se convierten en un conjunto de datos equilibrado de 17,805 imágenes después del aumento de datos del grupo de melanoma, obteniendo 8,903 imágenes de melanoma y 8,902 imágenes de no melanoma. Por lo tanto, el conjunto de datos final de las 17,805 imágenes, se divide aleatoriamente en 3 conjuntos de datos que contiene 2 clases (Fig. 2):

- **Entrenamiento:** 10,682 imágenes (60 %). Melanoma: 5,341 y No Melanoma: 5,341.
- **Validación:** 3,562 imágenes (20 %). Melanoma: 1,781 y No Melanoma: 1,781.
- **Prueba:** 3,561 imágenes (20 %). Melanoma: 1,781 y No Melanoma: 1,780.

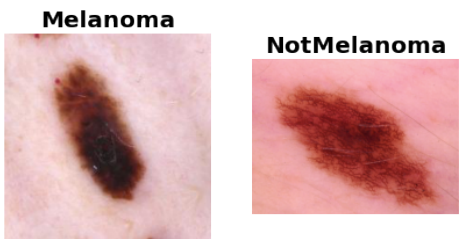


Figura 2: Muestra de una imagen de la clase de Melanoma y No Melanoma. Fuente: Fotos de [11] y acomodo propio.

### 3.2. Metodología

#### 3.2.1. Pre-procesamiento

Antes de pasar al entrenamiento del modelo con las imágenes desmatoscópicas, se aplicó un pre-procesamiento con el fin de reducir el efecto del ruido y patrones desequilibrados en el conjunto de datos que tienden a disminuir la capacidad de las CNN para aprender funciones importantes [7]. Primero, todas las imágenes se normalizan para escalar los valores de intensidad de píxeles de  $[0, 255]$  a  $[0, 1]$ . Después, se aplicó un aumento de datos para tener una mayor variedad de imágenes y que el modelo pueda diferenciar entre una clase y otra. Las técnicas de aumento aplicadas son el giro horizontal aleatorio, giro vertical aleatorio y una rotación aleatoria de  $45^\circ$ . En la Fig. 3 se puede observar la transformación de la imagen original con las técnicas aplicadas.

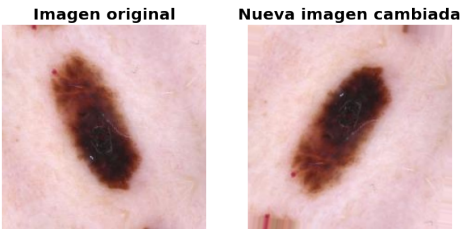


Figura 3: Muestra de la transformación de una imagen aplicando técnica de aumento de datos. Fuente: Fotos de [11] y acomodo propio.

Una diferencia marcada dentro del conjunto de datos son las diferentes dimensiones de las imágenes (Fig. 4), en donde, el promedio en anchura es de 435.26 píxeles y de altura 350.98 píxeles. Entonces, se decidió re-dimensionar las imágenes a un tamaño de 400x400 píxeles con la finalidad de conservar gran parte de la información, calidad, detalle en las imágenes.

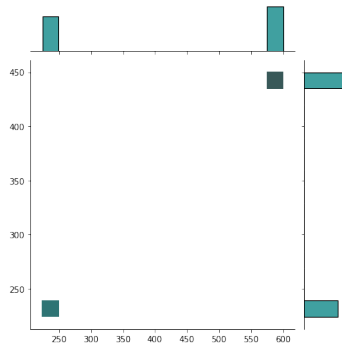


Figura 4: Distribución de la anchura y altura de las imágenes del conjunto de datos. Fuente: Imagen de elaboración propia.

Se eligió trabajar con las imágenes en RGB, ya que al cambiarlas en escala de grises no se obtenían buenos resultados. Por último, las etiquetas de cada clase se codificaron como 0 para No Melanoma y 1 para Melanoma.

### 3.2.2. Arquitecturas de Transfer Learning

Dentro de la literatura, se encuentran diversos estudios que muestran el desarrollo y la implementación de métodos precisos mediante el empleo de algoritmos de aprendizaje profundo. Estos desempeñan un papel importante en el diagnóstico de diversas enfermedades como el cáncer, la neumonía, el Alzheimer, entre otros [7]. Para este caso de clasificación de Melanoma en imágenes dermatoscópicas, el aprendizaje profundo resulta ser una herramienta muy favorable por sus resultados obtenidos que se hablará más adelante.

El Transfer Learning, es una metodología de aprendizaje profundo en la que un modelo que se ha entrenado para una tarea se utiliza como punto de partida para realizar otra tarea similar [9]. Estas arquitecturas son entrenadas a partir de un conjunto de datos llamado ImageNet, el cual contiene más de un millón de imágenes y una clasificación de mil clases.

Dada la anterior información, la motivación de este trabajo, es utilizar 5 arquitecturas de Transfer Learning para llevar a cabo la clasificación de las imágenes dermatoscópicas con la finalidad de comparar los resultados obtenidos y saber cuál de estos tuvo un mejor desempeño.

- **ResNet50 V2:** La arquitectura ResNet utiliza las funciones de cuello de botella para mejorar la eficiencia. Tiene 5 etapas, donde cada etapa tiene bloques de convolución e identidad. Trajo la idea de omitir conexiones para reducir el gradiente de fuga y el problema del gradiente de explosión que permiten que la red sea muy profunda [7, 13].
- **Xception:** La arquitectura Xception tiene 36 capas convolucionales para extraer características importantes y está inspirado en Inception Chollet, en el que los módulos de Inception se reemplazan con convoluciones separables en profundidad que consisten en una convolución en profundidad [5].
- **VGG16:** VGGNet utiliza filtros más pequeños de  $3 \times 3$  para proporcionar una mejor extracción de características de las imágenes. El uso de filtros mucho más pequeños para aumentar la profundidad de la red en lugar de su ancho juega un papel fundamental para obtener un mayor rendimiento. Hay dos versiones de esta arquitectura: VGG16 y VGG19 con diferentes profundidades y capas [7, 17].
- **MobileNet V2:** La arquitectura de MobileNetV2 contiene la capa inicial de convolución completa con 32 filtros, seguida de 19 capas de cuello de botella residuales. Usa ReLU6 como la no linealidad debido a su robustez cuando se usa con computación de baja precisión. Siempre usa el tamaño de kernel  $3 \times 3$  como es estándar para las redes modernas, y utilizamos el abandono y la normalización por lotes durante el entrenamiento [15].
- **Inception V3:** Presentado por primera vez como GoogLeNet y luego llamado Inception V1. Fue modificado varias veces desde entonces. Con la normalización por lotes, se denominó Inception V2 y con la idea de factorización adicional se denominó Inception V3. Realiza varias mejoras, incluido el uso de suavizado de etiquetas, convoluciones factorizadas de  $7 \times 7$  y el uso de un clasificador auxiliar para propagar información de etiquetas más abajo en la red [7, 19].

Como se mencionó, estas arquitecturas están diseñadas para clasificar 1,000 clases, por lo que se tiene que modificar la capa de salida para poder clasificar entre 2 clases. A continuación se describen cada una de las capas agregadas a las arquitecturas (Fig. 4):

1. El **Functional**, le corresponde a la arquitectura de Transfer Learning que se está utilizando.
2. El **Flatten**, le corresponde al aplanamiento de los mapas de características agrupadas en una columna [18].

3. En el primer **Dense**, se encuentran 256 neuronas conectadas con la capa anterior [20], que contienen una función de activación (**Activation**) ReLU. También, en esta primera capa Dense contiene la función **BactchNormalization** que aplica una transformación que mantiene la media de salida cercana a 0 y la desviación estándar de salida cercana a 1. A su vez un **Dropout** para evitar el sobreajuste en los datos de entrenamiento se apagan aleatoriamente el 50 % de las neuronas.
4. La última capa Dense contiene solo una neurona que será la encargada de clasificar las imágenes con ayuda de la función de activación Sigmoid que da como resultado valores que están entre 0 y 1. Como se está trabajando con una categoría binaria hay que determinar la etiqueta que le corresponde. Entonces, se define un umbral de clasificación de 0.5, donde los valores mayores al umbral se le asigna la etiqueta de 1 - Melanoma y menores o iguales al umbral se le asigna la etiqueta de 0 - No Melanoma.

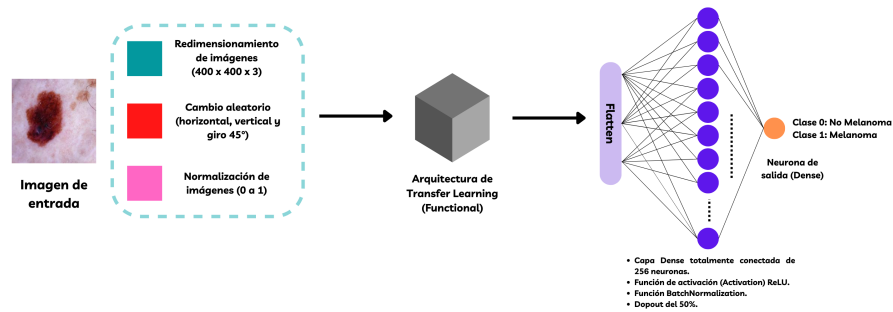


Figura 5: Diagrama de las arquitecturas de Transfer Learning modificadas. Fuente: Imagen de elaboración propia.

### 3.2.3. Métricas, función de perdida y optimizador

Si bien la preparación de los datos y el entrenamiento del modelo es un punto clave, también lo es el medir el rendimiento de este modelo entrenado. Entonces, se utilizará las siguientes métricas [2]:

- **Matriz de confusión:** Es una tabla que describe el desempeño de un modelo de clasificación en un conjunto de datos de prueba cuyos valores verdaderos son conocidos. Esta matriz de confusión puede ser utilizada para estimar otras métricas como en la Fig. 6[8].

En donde, **Verdadero Negativo (TN):** El valor real es negativo y la prueba predijo que el resultado era negativo.

**Verdadero Positivo (TP):** El valor real es positivo y la prueba predijo que era positivo.

**Falso Negativo (FN):** El valor real es positivo, y la prueba predijo que el resultado es negativo.

**Falso Positivo (FP):** El valor real es negativo, y la prueba predijo que el resultado es positivo.

- **Accuracy (Exactitud):** Es la proporción de resultados verdaderos (Verdaderos Positivos como Verdaderos Negativos) dividido entre el número total de casos examinados (Verdaderos Positivos, Falsos Positivos, Verdaderos Negativos, Falsos Negativos). Sin embargo, una de sus desventajas es que no funciona bien con clases desequilibradas[8].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision (Precisión):** Es la proporción de Verdaderos Positivos dividido entre los identificados como positivos (Verdaderos Positivos y Falsos Positivos). Sin embargo, una de sus desventajas es que no funciona bien con clases desequilibradas[8].

$$Precision = \frac{TP}{TP + FP}$$

- **Recall (Sensibilidad):** Es la proporción de Verdaderos Positivos que fueron correctamente identificados por el algoritmo entre el total de Verdaderos Positivos (Verdaderos Positivos y Falsos Negativos). Representa la fracción de Verdaderos Positivos[8].

$$Recall = \frac{TP}{TP + FN}$$

- **Specificidad (Especificidad):** Es la proporción de Verdaderos Negativos que fueron correctamente identificados por el algoritmo entre el total de Verdaderos Negativos (Verdaderos Negativos y Falsos Positivos). Representa la fracción de Verdaderos Negativos[8].

$$Specificidad = \frac{TN}{TN + FP}$$

- **F1 Score (Puntaje F1):** Asume que nos importa de igual forma la precisión y la exhaustividad[8].

$$F1 = 2\left(\frac{Precision * recall}{Precision + recall}\right)$$

- **AUC (Área Bajo la Curva):** Este puntaje nos da una buena idea de qué tan bien funciona el modelo.[8].

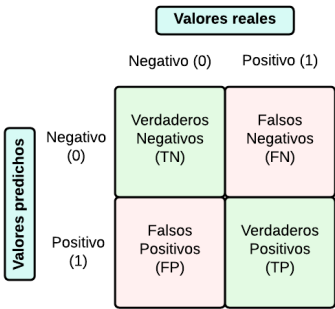


Figura 6: Representación de la matriz de confusión binaria. Fuente: Imagen de elaboración propia.

Por otra parte, como función de perdida se utilizó la entropía cruzada binaria, ya que penaliza la probabilidad en función de qué tan lejos está del valor esperado real. La penalización es de naturaleza logarítmica, lo que genera una puntuación grande para las diferencias grandes cercanas a 1 y una puntuación pequeña para las diferencias pequeñas que tienden a 0 [6]. También, como optimizador se utilizó el optimizador de Adam añadiendo la función ReduceLROnPlateau para reducir la tasa de aprendizaje cuando una métrica ha dejado de mejorar y la función EarlyStopping para dejar de entrenar el modelo cuando una métrica haya dejado de mejorar.

## 4. Resultados

En esta sección se muestras los resultados obtenidos de cada una de las arquitecturas propuestas para la clasificación de melanoma en imágenes dermatoscópicas los cuales se pueden observar en la Tab. 1 y Tab. 2. En el conjunto de entrenamiento, todas las arquitecturas obtuvieron un excelente desempeño encabezando Xception y MobileNet V2 con un 95.57 % y 95.18 % respectivamente. En el conjunto de validación de igual forma se obtuvieron resultados favorables encabezando ResNet50 V2 e Inception V3 con 94.75 % y 94.92 % respectivamente.

| Nombre de la arquitectura | Entrenamiento:<br>Exactitud<br>(Accuracy) | Validación:<br>Exactitud<br>(Accuracy) |
|---------------------------|---|--|
| ResNet50 V2               | 94.82 %                                   | <b>94.75 %</b>                         |
| Xception                  | <b>95.57 %</b>                            | 93.04 %                                |
| VGG16                     | 93.33 %                                   | 94.05 %                                |
| MobileNet V2              | <b>95.18 %</b>                            | 92.17 %                                |
| Inception V3              | 95.17 %                                   | <b>94.92 %</b>                         |

Tabla 1: Comparación de las arquitecturas utilizando diferentes métricas con el conjunto de entrenamiento y validación.

De igual forma cada una de las arquitectura mostró un gran desempeño al momento de clasificar si era melanoma o no con el conjunto de prueba. Las arquitecturas ResNet50 V2, Xception y VGG16 mostraron un mejor desempeño.

| Nombre de la arquitectura | Exactitud<br>(Accuracy) | Precisión<br>(Presicion) | Sensibilidad<br>(Recall) | Puntaje F1<br>(Score F1) |
|---------------------------|-------------------------|--------------------------|--------------------------|--------------------------|
| ResNet50 V2               | <b>93.57 %</b>          | <b>93.64 %</b>           | 93.49 %                  | <b>93.57 %</b>           |
| Xception                  | 92.78 %                 | 89.40 %                  | <b>97.08 %</b>           | <b>93.08 %</b>           |
| VGG16                     | <b>93.15 %</b>          | <b>98.73 %</b>           | 87.42 %                  | 92.73 %                  |
| MobileNet V2              | 90.90 %                 | 85.98 %                  | <b>97.75 %</b>           | 91.49 %                  |
| Inception V3              | 92.64 %                 | 89.25 %                  | 96.97 %                  | 92.95 %                  |

Tabla 2: Comparación de las arquitecturas utilizando diferentes métricas con el conjunto de prueba.

Tomando en cuenta que la mejor arquitectura fue ResNet50 V2, en la Fig. 7 se puede observar la evolución de cada una de las métricas utilizadas para determinar el desempeño de la arquitectura en el conjunto de entrenamiento y validación. Se utilizó 50 épocas para entrenar todas las arquitecturas, pero

como se observa en cada una de las gráficas (Fig. 7), en el caso de ResNet50 V2 no llegó a las 50 y se detuvo en la época 40. Esto se debe a las funciones antes mencionadas: ReduceLROnPlateau y EarlyStopping.

En general, las gráficas de la Fig. 7 tienen un excelente ajuste y este se identifica por una pérdida de entrenamiento y validación (Fig. 7d) que disminuye hasta un punto de estabilidad con una brecha mínima entre los dos valores de pérdida finales. La pérdida del modelo casi siempre será menor en el entrenamiento que en el de validación y se esperaría una pequeña brecha entre estas dos conocida como brecha de generalización [3].

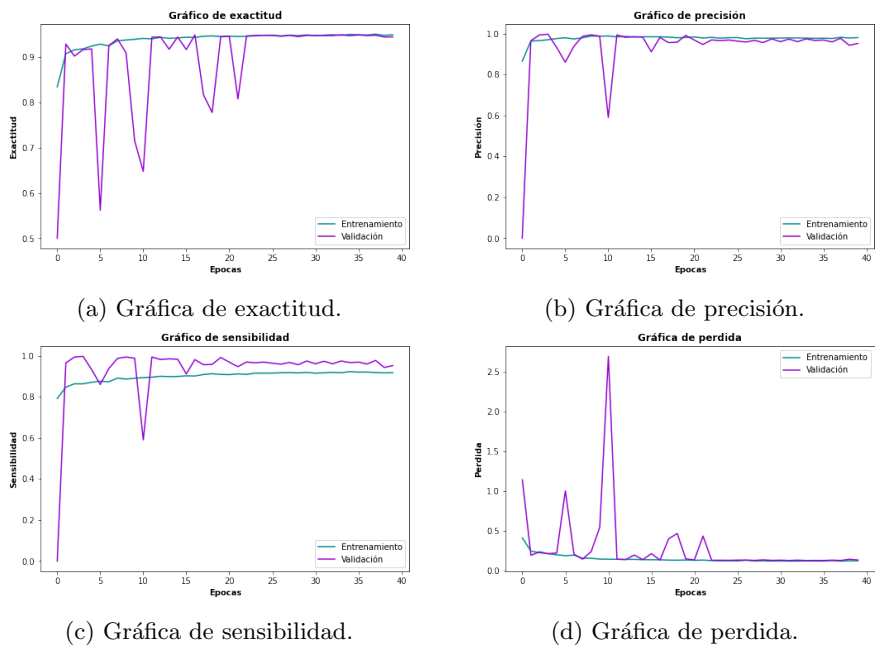


Figura 7: Gráficos de las métricas utilizadas de la arquitectura ResNet50 V2. Fuente: Imagen de elaboración propia.

Una vez entrenado el modelo se procede a evaluarlo con el conjunto de prueba, y como se había mencionado, de acuerdo al umbral de clasificación se construye la matriz de confusión. Esta matriz revelará qué es lo que esta clasificando bien y mal entre "Melanoma" y "No Melanoma", como se puede observar en la Fig. 8 se encuentra la matriz de confusión de cada una de las arquitecturas de Transfer Learning utilizadas. En la matriz de confusión de ResNet50 V2 (Fig. 8a) existe un balanceo entre los falsos negativos y falsos positivos, mientras que en los demás la confusión cae en los falsos positivos a excepción de VGG16 que hay más falsos negativos.

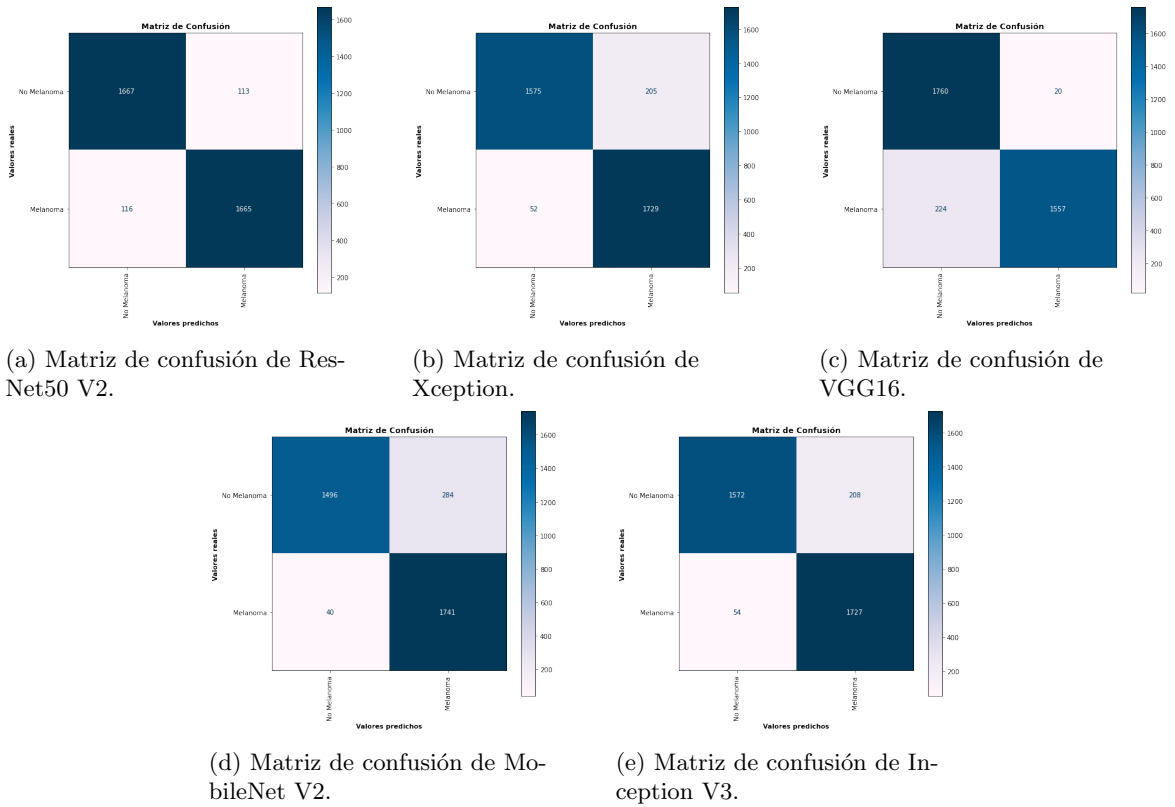


Figura 8: Matriz de confusión de cada una de las arquitecturas utilizadas. Fuente: Imagen de elaboración propia.



Con ayuda de la sensibilidad y 1-especificidad se puede construir la curva ROC (Característica Operativa del Receptor), es un gráfico que muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación y 1-especificidad es la tasa de falsos positivos. En la Fig. 9 se puede observar la curva ROC de cada una de las arquitecturas de Transfer Learning utilizadas, en el cual, ResNet50 V2 es la que destaca por su AUC de 93.57 % (Fig. 9a) seguido de VGG16 (Fig. 9c).

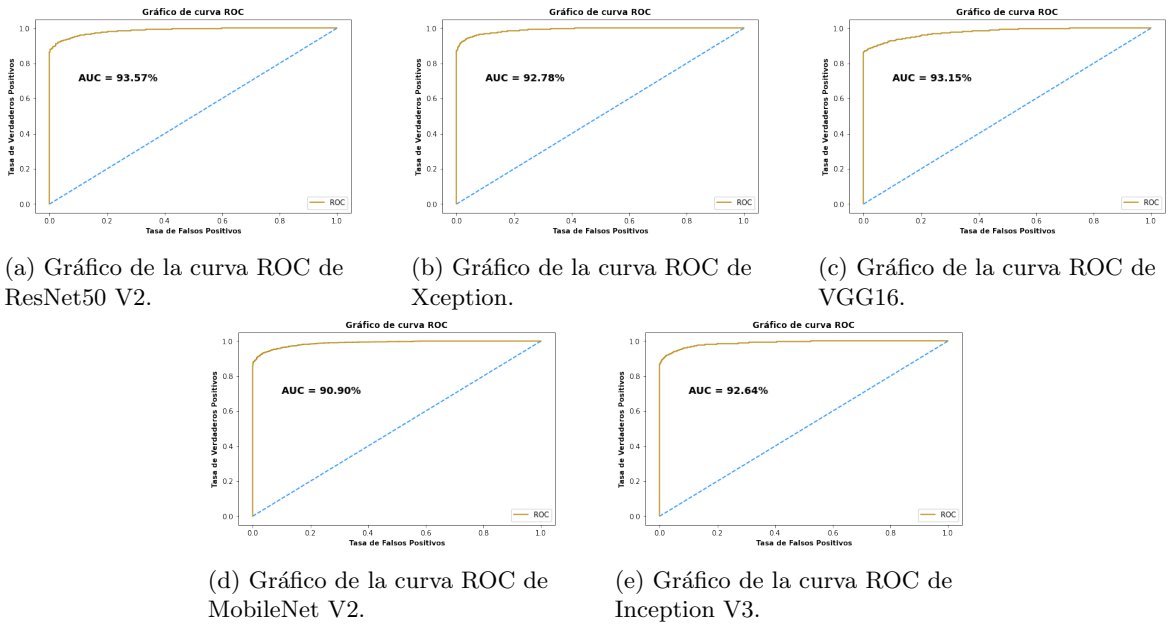


Figura 9: Gráfico de la curva ROC de cada una de las arquitecturas utilizadas. Fuente: Imagen de elaboración propia.

Utilizando la arquitectura ResNet50 V2 se realizó una predicción de varias imágenes (Fig. 10) y se puede observar la clasificación real a la que pertenece, la clasificación predicha por el modelo y la probabilidad para decidir si es Melanoma o No Melanoma.

**Predicciones del modelo (Verde: Correcto, Rojo: Incorrecto)**

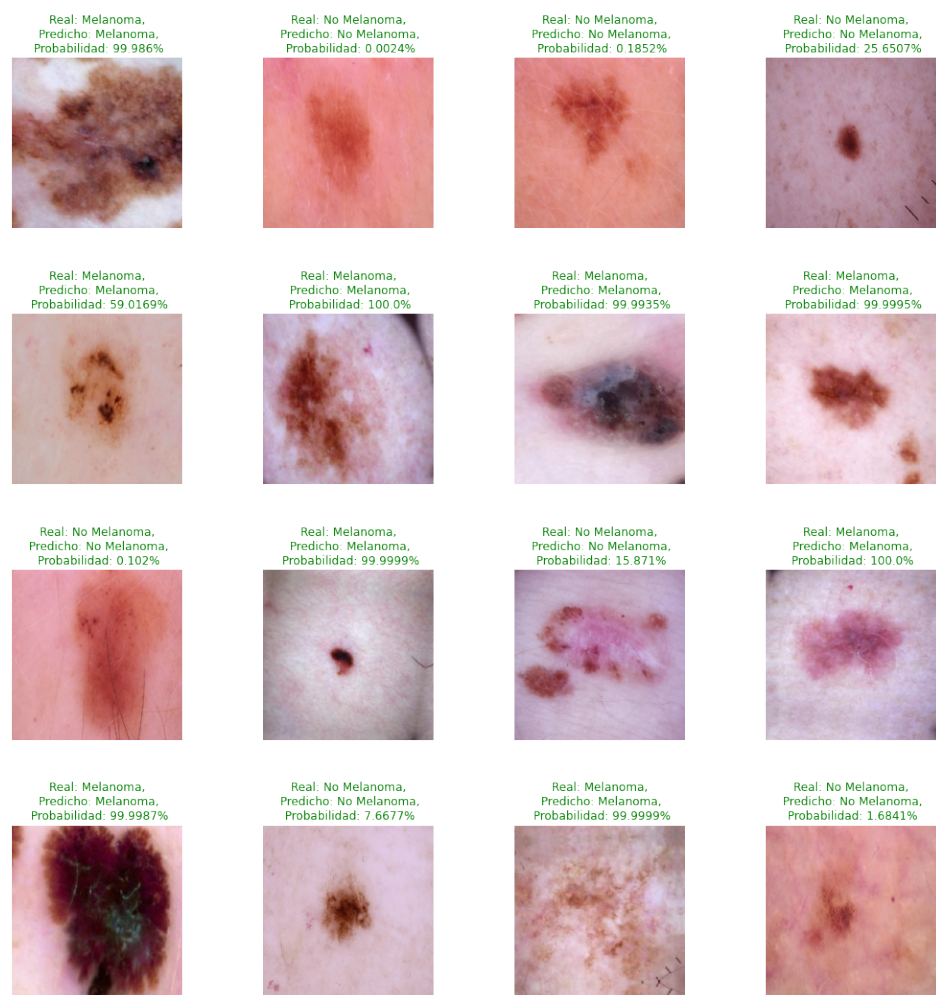


Figura 10: Predicción de varias imágenes. Fuente: Fotos de [11] y acomodo propio.

## 5. Discusión

De acuerdo a los resultados obtenidos en la sección anterior y comparando los resultados que se obtuvieron en el estado del arte utilizando metodologías similares, se puede concluir que son muy similares. En el artículo [13] las arquitecturas ResNet50, Xception, VGG16 e Inception V3 obtuvieron de exactitud (accuracy) 95.79 %, 96.10 %, 94.83 % y 95.48 % respectivamente. Este mismo menciona que ResNet50 y Xception fueron los de mejor desempeño.

En el trabajo [14] propuso arquitecturas como Xception, Inception, ResNet50 V2, DenseNet121 y DenseNet201 logrando una exactitud de 95,17 %, 95,17 %, 94,48 % y 95,86 %, respectivamente. De igual forma en el artículo [7], se examinó el desempeño de AlexNet, ResNet50, VGGNet16, VGGNet19 y Xception obteniendo 80.45 %, 92.08 %, 88.36 %, 88.70 %, 90.30 % respectivamente. Este último, eligió a la arquitectura ResNet50 como la mejor.

En este trabajo, la arquitectura ResNet50 V2 dio mejores resultados y de acuerdo con el estado del arte, también se destacaba de las demás seguido de Xception.

## 6. Conclusión

En este trabajo se presentó una comparación entre distintas arquitecturas para la detección de melanoma en imágenes dermatoscópicas demostrando que el empleo de métodos de aprendizaje profundo pueden ser una solución viable para resolver problemas médicos.

La investigación y construcción de diferentes métodos nos da una vista más amplia sobre que metodología hay que elegir para obtener los mejores resultados. En este caso, al emplear una técnica para cambiar las imágenes en nuestro conjunto de datos tuvo un gran impacto en los resultados y en la mejora de la precisión del modelo.

En trabajos futuros, se busca tener acceso a un amplio conjunto de datos, en donde no solo se detecte melanoma sino otras enfermedades de la piel. También, construir una CNN propia como en los artículos [1, 12] y compararlos con las arquitecturas de Transfer Learning, aplicando y/o agregando más técnicas de aumento de datos para tener más variedad.

## Referencias

- [1] Aya Abu Ali y Hasan Al-Marzouqi. «Melanoma detection using regular convolutional neural networks». En: *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. IEEE. 2017, págs. 1-5.
- [2] Ricardo Borja-Robalino, Antonio Monleón-Getino y José Rodellar. «Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning». En: *Revista Ibérica de Sistemas e Tecnologías de Informação E30* (2020), págs. 184-196.
- [3] Jason Brownlee. *Función de pérdida de entropía cruzada*. URL: <https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/>. (Acceso: 07-07-2022).
- [4] Instituto Nacional del Cáncer. *Neoplasia*. URL: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/neoplasia>. (Acceso: 13-07-2022).
- [5] François Chollet. «Xception: Deep learning with depthwise separable convolutions». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, págs. 1251-1258.
- [6] ICHI.PRO. *Función de pérdida de entropía cruzada*. URL: <https://ichi.pro/es/funcion-de-perdida-de-entropia-cruzada-267783942726718>. (Acceso: 07-07-2022).
- [7] Sara Hosseinzadeh Kassani y Peyman Hosseinzadeh Kassani. «A comparative study of deep learning architectures on melanoma detection». En: *Tissue and Cell* 58 (2019), págs. 76-83.
- [8] AprendeIA con Ligdi Gonzalez. *MÉTRICAS DE EVALUACIÓN MODELOS DE CLASIFICACIÓN SCIKIT LEARN — #35 Curso Machine Learning con Python*. URL: <https://www.youtube.com/watch?v=K5PNrX694HQ>. (Acceso: 07-07-2022).
- [9] MathWorks. *Transfer Learning*. URL: <https://la.mathworks.com/discovery/transfer-learning.html>. (Acceso: 07-07-2022).
- [10] Alexander Scarlat MD. *Machine Learning Primer for Clinicians-Part 12*. URL: <https://histalk2.com/2019/01/16/machine-learning-primer-for-clinicians-part-12-2/>. (Acceso: 07-07-2022).
- [11] Alexander Scarlat MD. *melanoma*. URL: <https://www.kaggle.com/datasets/drscarlat/melanoma>. (Acceso: 07-07-2022).
- [12] Soumen Mukherjee, Arunabha Adhikari y Madhusudan Roy. «Malignant melanoma classification using cross-platform dataset with deep learning CNN architecture». En: *Recent trends in signal and image processing*. Springer, 2019, págs. 31-41.
- [13] Md Fazle Rasul, Nahin Kumar Dey y MMA Hashem. «A comparative study of neural network architectures for lesion segmentation and melanoma detection». En: *2020 IEEE Region 10 Symposium (TENSYP)*. IEEE. 2020, págs. 1572-1575.
- [14] Rehan Raza y col. «Melanoma Classification from Dermoscopy Images Using Ensemble of Convolutional Neural Networks». En: *Mathematics* 10.1 (2022). ISSN: 2227-7390. DOI: 10.3390/math10010026. URL: <https://www.mdpi.com/2227-7390/10/1/26>.
- [15] Mark Sandler y col. «Mobilenetv2: Inverted residuals and linear bottlenecks». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, págs. 4510-4520.
- [16] Dirk Schadendorf y col. «Melanoma». En: *Nature reviews Disease primers* 1.1 (2015), págs. 1-20.
- [17] Karen Simonyan y Andrew Zisserman. «Very deep convolutional networks for large-scale image recognition». En: *arXiv preprint arXiv:1409.1556* (2014).
- [18] SuperDataScience. *Convolutional Neural Networks (CNN): Step 3 - Flattening*. URL: <https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-3-flattening>. (Acceso: 07-07-2022).
- [19] Christian Szegedy y col. «Rethinking the inception architecture for computer vision». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 2818-2826.
- [20] Yugesh Verma. *A Complete Understanding of Dense Layers in Neural Networks*. URL: <https://analyticsindiamag.com/a-complete-understanding-of-dense-layers-in-neural-networks/>. (Acceso: 07-07-2022).
- [21] Medical University of Vienna. *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*. URL: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T>. (Acceso: 07-07-2022).