

# Desarrollo de un Modelo de Inteligencia Artificial para el Diagnóstico de Melanoma

Marc Pérez Guerrero

Estudiante en Grado de Ingeniería Informática de la EPSEVG-UPC

## Resumen

En este Trabajo de Final de Grado presento el desarrollo de un modelo de inteligencia artificial (IA) capaz de detectar automáticamente melanomas a partir del análisis de imágenes dermatológicas. El objetivo es ofrecer una herramienta adicional que facilite a los especialistas la detección temprana de este tipo de cáncer. Para ello, he implementado y comparado dos arquitecturas diferentes: Redes Neuronales Convolucionales (CNN) y Vision Transformers (ViT), entrenadas con bases de datos especializadas como HAM10000, ISIC y BCN20000.

Durante el desarrollo, he aplicado diversas técnicas avanzadas, incluyendo preprocesamiento de imágenes, aumento de datos (data augmentation) y optimización de hiperparámetros mediante Optuna con el algoritmo TPE (Tree-Structured Parzen Estimator). Tras evaluar ambas arquitecturas, he seleccionado finalmente un modelo basado en Vision Transformers, debido a su mayor rendimiento en términos de precisión, sensibilidad y especificidad, alcanzando métricas superiores al 97%.

El modelo desarrollado puede ser considerado como un componente inicial que, en futuras investigaciones, podría ser ampliado y optimizado para su incorporación en entornos clínicos reales.

## 1. Introducción

El melanoma es uno de los tipos de cáncer de piel más agresivos, y su diagnóstico temprano es esencial para mejorar el pronóstico de los pacientes. La detección precoz sigue siendo compleja, principalmente debido a la gran variabilidad en la apariencia visual de las lesiones cutáneas. Actualmente, el diagnóstico inicial del melanoma recae principalmente en el dermatólogo, que mediante técnicas visuales como la dermatoscopia evalúa si es necesario realizar pruebas complementarias, tales como biopsias [1] (“Cáncer de piel”).

La Inteligencia Artificial (IA), en especial mediante técnicas de aprendizaje profundo, ha mostrado en los últimos años resultados prometedores en el ámbito del diagnóstico médico automatizado. Entre estas técnicas destacan las Redes Neuronales Convolucionales (CNN), reconocidas por su efectividad en tareas de visión por computadora. Más recientemente, los modelos basados en Vision Transformers (ViT) han demostrado una gran capacidad para capturar relaciones complejas en imágenes, alcanzando resultados aún más prometedores en diferentes campos médicos [2] (“Avance Redes Neuronales En Aplicaciones Médicas”).

La motivación para realizar este Trabajo de Final de Grado surge de mi interés por desarrollar una herramienta basada en

IA que pueda asistir a los dermatólogos en la clasificación automática de lesiones de melanoma. Para alcanzar este objetivo, he decidido implementar dos arquitecturas diferentes (CNN y ViT) de forma independiente, entrenándolas y comparándolas con bases de datos especializadas como HAM10000 [3] (“The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions”), ISIC [4] (ISIC), BCN20000 [5] (BCN20000) y Kaggle [6] (“Melanoma Skin Cancer Dataset of 10000 Images”). Para optimizar aún más el rendimiento de estos modelos, he aplicado técnicas avanzadas de preprocesamiento, aumento de datos (data augmentation) y optimización de hiperparámetros mediante Optuna [7] (“Optuna: A hyperparameter optimization framework”), empleando el algoritmo TPE (Tree-Structured Parzen Estimator).

Finalmente, tras evaluar rigurosamente ambas arquitecturas, he seleccionado como modelo definitivo el Vision Transformer, que ha mostrado un rendimiento superior en precisión, sensibilidad y especificidad. Este trabajo sienta las bases para futuras investigaciones y potenciales aplicaciones clínicas reales, ofreciendo una herramienta robusta capaz de apoyar el diagnóstico temprano del melanoma y contribuir a la mejora de las tasas de supervivencia.

## 2. Estado del Arte en el Diagnóstico de Melanoma mediante Inteligencia Artificial

El diagnóstico de melanoma ha sido tradicionalmente una tarea compleja para los dermatólogos, basada en la inspección visual y técnicas como la dermatoscopia. Sin embargo, este enfoque presenta limitaciones debido a la subjetividad en la evaluación y la necesidad de experiencia clínica para interpretar correctamente las lesiones cutáneas.

Inicialmente, los sistemas computacionales de ayuda al diagnóstico se basaban en técnicas de procesamiento de imágenes convencionales, como la segmentación de lesiones y el análisis manual de características. Estos métodos, aunque útiles, tenían una capacidad de generalización limitada y dependían en gran medida de la intervención humana.

Con el avance del aprendizaje profundo, los modelos de redes neuronales convolucionales (CNN) han revolucionado el análisis de imágenes médicas. Estudios recientes han demostrado que las CNN pueden alcanzar un rendimiento comparable al de dermatólogos expertos en la clasificación de lesiones cutáneas, mejorando la precisión y reduciendo el tiempo de diagnóstico. Bases de datos como HAM10000, ISIC y BCN20000 han sido ampliamente utilizadas para entrenar estos modelos, proporcionando conjuntos de

imágenes etiquetadas que permiten mejorar la capacidad de detección del melanoma.

Más recientemente, los Vision Transformers (ViT) han surgido como una alternativa a las CNN en tareas de visión por computadora. A diferencia de las redes convolucionales, los ViT utilizan un mecanismo de autoatención que permite capturar relaciones globales en las imágenes, en lugar de centrarse en regiones locales. Este enfoque ha demostrado ser eficaz en diversas aplicaciones médicas, incluido el diagnóstico de enfermedades dermatológicas, al permitir una mejor identificación de patrones visuales complejos.

A pesar de estos avances, la implementación de modelos de inteligencia artificial en el ámbito clínico enfrenta desafíos importantes, como el desbalance de clases en las bases de datos, la variabilidad en la calidad y el tipo de imágenes, y la necesidad de interpretabilidad en los resultados para facilitar su uso por parte de los profesionales de la salud. Además, aunque estos modelos han mostrado un alto rendimiento en entornos controlados, su generalización a imágenes obtenidas en condiciones reales sigue siendo un reto.

### 3. Características Clínicas del Melanoma

El melanoma es una de las formas más agresivas de cáncer de piel y se origina en los melanocitos, las células responsables de la producción de melanina. Su detección temprana es fundamental, ya que las tasas de supervivencia disminuyen drásticamente cuando la enfermedad progresa a estadios avanzados.

El diagnóstico clínico del melanoma se basa en la identificación de patrones visuales en las lesiones cutáneas. Uno de los métodos más utilizados por los dermatólogos es la regla ABCDE, que permite diferenciar lesiones benignas de posibles melanomas:

**Asimetría (A):** Los melanomas suelen presentar formas irregulares, mientras que las lesiones benignas tienden a ser más simétricas.

**Bordes (B):** Los bordes de los melanomas suelen ser irregulares, mal definidos o difusos, a diferencia de los lunares benignos, cuyos contornos son más uniformes.

**Color (C):** La presencia de múltiples tonalidades en una misma lesión (negro, marrón, azul, rojo o blanco) es un indicador de melanoma, mientras que las lesiones benignas suelen tener un color uniforme.

**Diámetro (D):** Las lesiones malignas suelen ser mayores a 6 mm, aunque esta no es una regla definitiva.

**Evolución (E):** Un cambio en la forma, tamaño, color o textura de una lesión a lo largo del tiempo es un factor clave para su diagnóstico.

Además de la inspección visual, técnicas como la dermatoscopia permiten una evaluación más detallada de las lesiones cutáneas, facilitando la identificación de patrones pigmentarios y estructuras internas que pueden ser indicativos de malignidad [8] (“Qué buscar: ABCDE de melanoma”).

En el desarrollo del modelo de este trabajo, se han tenido en cuenta principalmente los criterios A (Asimetría), B (Bordes) y C (Color), ya que los otros dos (Diámetro y Evolución) requieren información adicional que no siempre está

disponible en las bases de datos de imágenes. La falta de metadatos sobre el tamaño real de las lesiones y la ausencia de seguimiento temporal de las imágenes dificulta la incorporación de estos factores en la detección automatizada.

### 4. Arquitecturas Exploradas: Redes Neuronales Convolucionales y Vision Transformers

El desarrollo de un modelo de inteligencia artificial para el diagnóstico de melanoma requiere elegir una arquitectura capaz de extraer y procesar características relevantes de las imágenes dermatológicas. En este trabajo, se han explorado dos enfoques principales: Redes Neuronales Convolucionales (CNN) y Vision Transformers (ViT), cada una con sus ventajas y limitaciones.

#### Redes Neuronales Convolucionales (CNN):

Las Redes Neuronales Convolucionales (CNN) han sido ampliamente utilizadas en el análisis de imágenes médicas debido a su capacidad para extraer patrones visuales relevantes mediante el uso de filtros convolucionales. Su estructura básica incluye:

- Capas convolucionales: Aplican filtros a la imagen para detectar bordes, texturas y otros patrones visuales.
- Capas de pooling: Reducen la dimensionalidad de los mapas de características, preservando la información más relevante y mejorando la eficiencia computacional.
- Capas totalmente conectadas: Al final de la red, estas capas interpretan las características extraídas y generan una predicción final.

Las CNN han demostrado un alto rendimiento en la clasificación de imágenes dermatológicas. Sin embargo, su principal limitación es que operan sobre regiones locales de la imagen, lo que puede dificultar la captura de relaciones globales entre distintas partes de una lesión cutánea.

#### Vision Transformers (ViT)

Los Vision Transformers (ViT) son una arquitectura más reciente que ha demostrado un gran potencial en tareas de visión por computadora. A diferencia de las CNN, los ViT dividen la imagen en pequeños bloques llamados patches, los cuales se procesan utilizando un mecanismo de autoatención. Este enfoque permite que el modelo aprenda relaciones entre diferentes regiones de la imagen de manera más efectiva.

El proceso de un ViT se compone de:

- División de la imagen en patches: La imagen se segmenta en pequeños fragmentos de tamaño fijo.
- Conversión a embeddings: Cada patch se convierte en un vector numérico que representa su información visual.
- Mecanismo de autoatención: Se calcula la relación entre todos los patches de la imagen para capturar dependencias globales.
- Clasificación final: Se genera una predicción basada en la representación global de la imagen.

Los ViT han demostrado ser altamente efectivos en tareas de visión por computadora cuando se entrenan con grandes volúmenes de datos. Sin embargo, requieren más recursos

computacionales que las CNN y pueden ser menos eficientes en entornos con datos limitados.

## 5. Bases de datos e imágenes

### Origen de las imágenes

Para desarrollar el modelo he utilizado imágenes procedentes de bases de datos especializadas ampliamente reconocidas en el ámbito dermatológico. Las fuentes seleccionadas han sido los conjuntos HAM10000, ISIC, BCN20000 y Kaggle, debido a su relevancia clínica, calidad técnica y aceptación en la comunidad científica.

Las bases de datos HAM10000 (Human Against Machine 10000) y Kaggle, ofrecen una colección extensa y representativa de imágenes dermatoscópicas de diferentes lesiones cutáneas, obtenidas en contextos clínicos reales y previamente etiquetadas por expertos dermatólogos. Por otro lado, ISIC (International Skin Imaging Collaboration) proporciona una amplia variedad de imágenes dermatológicas que destacan por su diversidad clínica y visual, lo que resulta fundamental para lograr un modelo robusto y generalizable. Finalmente, BCN20000 contiene imágenes dermatológicas obtenidas en el contexto clínico del Hospital Clínic de Barcelona, lo que añade un valor adicional de realismo y pertinencia a los datos utilizados.

La construcción del conjunto final de imágenes, denominada Database\_v2, ha pasado por diferentes etapas de evolución y mejora. Partiendo desde una versión inicial denominada Database\_v0, se ha ido refinando progresivamente hasta llegar a la versión definitiva Database\_v2. Este proceso evolutivo ha permitido incrementar la calidad general del conjunto mediante la eliminación de imágenes inadecuadas, la incorporación de nuevas imágenes para mejorar la generalización del modelo y aplicar más actualización de datos para representar mayores variabilidades en las imágenes.

La versión final de la base de datos (Database\_v2) utilizada en este trabajo es el resultado de dicho proceso iterativo, diseñado específicamente para optimizar el entrenamiento y validación del modelo, asegurando así resultados fiables y clínicamente relevantes.

### Organización y distribución de los datos

Para asegurar una correcta evaluación del rendimiento del modelo y facilitar el proceso de entrenamiento, he organizado los datos en tres subconjuntos específicos: entrenamiento, validación y test. La distribución ha sido elegida, con la finalidad de asegurar que el modelo generalice adecuadamente y evitar posibles problemas de sobreajuste.

La base de datos final, está compuesta por un total de 156.344 imágenes dermatológicas, distribuidas del siguiente modo:

- Conjunto de entrenamiento: 138.492 imágenes
  - 69.246 benignas
  - 69.246 malignas
- Conjunto de validación: 8.925 imágenes
  - 7.419 benignas
  - 1.506 malignas
- Conjunto de prueba: 8.927 imágenes
  - 7.420 benignas
  - 1.507 malignas

La división adoptada respeta la proporción del 60% para entrenamiento, 20% para validación y 20% para prueba, asegurando así una evaluación rigurosa y representativa del rendimiento del modelo desarrollado.

### Preprocesamiento de imágenes

Para garantizar la calidad e integridad del conjunto final de imágenes (Database\_v2), se aplicaron dos filtros específicos con el objetivo de eliminar posibles imágenes problemáticas, duplicadas o con defectos visuales importantes, siguiendo el proceso descrito a continuación:

- Eliminación de imágenes duplicadas: Se implementó un algoritmo basado en hashing perceptual (average\_hash de la librería imagehash) para identificar imágenes duplicadas o muy similares. Inicialmente, el algoritmo se configuró con un umbral alto para detectar imágenes con similitudes evidentes. Las imágenes identificadas fueron trasladadas a un directorio específico para su revisión manual, donde se verificaron una por una para confirmar la duplicidad antes de eliminarlas definitivamente. Posteriormente, el umbral fue ajustado progresivamente a valores menores para identificar duplicados más sutiles, asegurando así una limpieza exhaustiva del conjunto de datos.
- Filtro de imágenes con colores uniformes: Tras eliminar duplicados, se implementó un filtro específico para detectar imágenes completamente negras, blancas o de un solo tono uniforme, ya que estas no proporcionan información útil para la tarea de clasificación. Este filtro inicialmente identificaba únicamente imágenes completamente uniformes, pero gradualmente se ajustó para detectar también aquellas imágenes con colores casi uniformes. Todas las imágenes detectadas fueron inspeccionadas manualmente para decidir si aportaban información relevante o si debían descartarse definitivamente.

### Data augmentation

Se utilizó para incrementar la variabilidad del conjunto de imágenes, facilitando así un mejor entrenamiento y generalización del modelo. Las transformaciones aplicadas sobre las imágenes dermatológicas fueron las siguientes:

Rotación con Zoom:

Se aplicó una rotación aleatoria en un rango comprendido entre  $-45^\circ$  y  $45^\circ$ , combinada con un zoom aleatorio ajustado en un rango de 0.9x a 1.1x del tamaño original.

Ajuste de brillo:

Se modificó el brillo de las imágenes mediante la multiplicación aleatoria de la componente de luminancia en el espacio de color HSV por un factor entre 0.8 y 1.2.

Ajuste de contraste:

El contraste se ajustó alterando la componente de luminosidad en el espacio de color LAB por un factor aleatorio comprendido entre 0.8 y 1.2.

Volteo horizontal y vertical:

Se aplicó una inversión aleatoria de las imágenes en los ejes horizontal y vertical.

Desplazamiento aleatorio:

Se desplazaron aleatoriamente las imágenes en los ejes X e Y hasta un máximo del 20% del tamaño total de la imagen original.

Debido al fuerte desbalance entre las clases benignas y malignas en la base de datos original, se estableció una estrategia diferenciada para cada clase:

- Para la clase benigna, se generó una única imagen aumentada por cada imagen original, aplicando aleatoriamente una de las transformaciones mencionadas anteriormente.
- Para la clase maligna, dada la menor cantidad de imágenes originales, se generaron nueve imágenes aumentadas adicionales por cada imagen original, aplicando una transformación diferente para cada una.

Finalmente, para evitar un exceso de imágenes en alguna de las clases y mantener el equilibrio, se realizó un ajuste eliminando alguna de las imágenes generadas artificialmente para la clase mayoritaria, hasta alcanzar un balance exacto entre ambas clases.

De esta manera, tras aplicar las transformaciones mencionadas y realizar los ajustes necesarios, la base de datos final quedó constituida por un total de 156.344 imágenes, perfectamente balanceadas.

## 6. Evolución y desarrollo del modelo

A lo largo del desarrollo del proyecto se definieron diferentes series de modelos, cada una representando una arquitectura o configuración específica con cambios significativos respecto a la serie anterior. En total, generé cinco series distintas (desde la serie v0 hasta la serie v4), dentro de las cuales probé varios modelos con distintas configuraciones y enfoques, siempre orientados a mejorar progresivamente el rendimiento del modelo.

### Serie v0:

Esta primera iteración estuvo formada por un único modelo, con una arquitectura CNN sencilla. El objetivo de esta serie fue familiarizarme con el entrenamiento del modelo y comprender cómo interactuaban los distintos hiperparámetros y técnicas básicas de preprocesamiento. No se buscó optimizar específicamente su rendimiento, sino sentar las bases para futuras mejoras.

### Serie v1:

En esta segunda iteración, se aumentó la complejidad del modelo añadiendo capas adicionales y probando técnicas avanzadas, como la reducción dinámica de la tasa de aprendizaje (ReduceLROnPlateau) y la optimización automática de hiperparámetros mediante Optuna. Se probaron tres modelos distintos (v1.0, v1.1 y v1.2), obteniendo en esta última versión resultados significativamente mejores que la serie v0 en términos de precisión y generalización.

### Serie v2:

La tercera iteración se basó en un enfoque distinto, utilizando transferencia de aprendizaje con la arquitectura EfficientNetB0 preentrenada en ImageNet. El objetivo de esta serie era aprovechar las capacidades avanzadas de detección de patrones ofrecidas por EfficientNetB0. Se realizaron múltiples pruebas (desde el modelo v2.0 hasta el modelo v2.0.2), explorando ajustes como la descongelación de capas y optimizaciones específicas. El mejor modelo de esta serie, el v2.0.2, logró un rendimiento destacado en precisión (accuracy de 0.9721), aunque evidenció problemas de generalización a otros datasets distintos al original, lo que motivó continuar explorando nuevas opciones.

### Serie v3:

En la cuarta iteración, decidí explorar otras arquitecturas distintas para superar los problemas de generalización encontrados en la serie v2. Concretamente, probé modelos basados en ResNet y una combinación de ResNet con EfficientNetB0. Sin embargo, ninguno de estos modelos logró mejorar significativamente los resultados previos, por lo que se descartó esta serie y se decidió avanzar hacia un enfoque radicalmente distinto en la siguiente iteración.

### Serie v4 (donde pertenece el modelo final):

La quinta y última iteración representó un cambio completo respecto a todas las anteriores, adoptando un enfoque basado en Vision Transformers (ViT). Utilicé el modelo ViT-B/16 preentrenado en ImageNet, permitiendo el entrenamiento completo del modelo sin congelar capas. Dentro de esta serie probé cinco versiones diferentes del modelo (desde v4.0 hasta v4.4). Finalmente, seleccioné la versión v4.3 como modelo definitivo, ya que logró el mejor rendimiento de todos los modelos de este trabajo.

## 7. Modelo final

La versión definitiva del modelo, denominada v4.3, está basada en la arquitectura Vision Transformer (ViT-B/16), preentrenada en ImageNet. Este modelo se entrenó completamente, sin congelar ninguna de sus capas.

La elección y optimización específica de esta arquitectura se realizó mediante un proceso de optimización de hiperparámetros utilizando Optuna. Concretamente, se optimizaron los siguientes hiperparámetros:

- Número de unidades densas (dense\_units), optimizado dentro del rango [128–512].
- Tasa de dropout (dropout\_rate): optimizada dentro del rango [0.4–0.7].
- Tasa de aprendizaje (learning\_rate): optimizada en un intervalo logarítmico entre [1e-5–5e-4].
- Regularización L2 (l2\_rate): optimizada también en un intervalo logarítmico [1e-6–1e-2].

Además, el entrenamiento del modelo final utilizó la función de pérdida Focal Loss con el parámetro  $\alpha$  establecido en 0.75 y pesos específicos para cada clase debido al desbalance existente en los datos:

- Clase benigna: peso de 0.7
- Clase maligna: peso de 1.3

Finalmente, también se implementaron técnicas complementarias destinadas a mejorar la convergencia, optimizar recursos y guardar el modelo:

- EarlyStopping, para detener anticipadamente el entrenamiento si no se obtenían mejoras significativas en la métrica de validación durante varias épocas consecutivas.
- ReduceLROnPlateau, para reducir automáticamente la tasa de aprendizaje cuando el rendimiento del modelo dejaba de mejorar.
- ModelCheckpoint, para guardar automáticamente el mejor modelo obtenido en función de la métrica de validación.

### 8 Resultados

Para evaluar el rendimiento del modelo final v4.3, se utilizó el conjunto de prueba de Database\_v2, el cual consta de un total de 8.927 imágenes, distribuidas de la siguiente manera:

- 7.420 imágenes de lesiones benignas
- 1.507 imágenes de melanomas malignos

Las métricas obtenidas en esta evaluación fueron las siguientes:

**F1 Score:** 0.9755  
**Recall (Sensibilidad):** 0.9642  
**Especificidad:** 0.9974  
**Accuracy:** 0.9918  
**AUC-ROC:** 0.9965

Estos resultados reflejan que el modelo es altamente preciso en la clasificación de imágenes dermatológicas, con un equilibrio óptimo entre sensibilidad y especificidad. La alta precisión y el elevado valor de Recall confirman que el modelo es capaz de detectar correctamente la mayoría de los casos de melanoma, minimizando la posibilidad de falsos negativos.

Cabe destacar que este conjunto de test no se utilizó en ninguna fase de entrenamiento ni ajuste del modelo, por lo que estas métricas reflejan el rendimiento real del modelo en imágenes completamente nuevas.

En este conjunto el modelo esta seguro de sus predicciones, esto lo podemos ver reflejado en la siguiente figura.

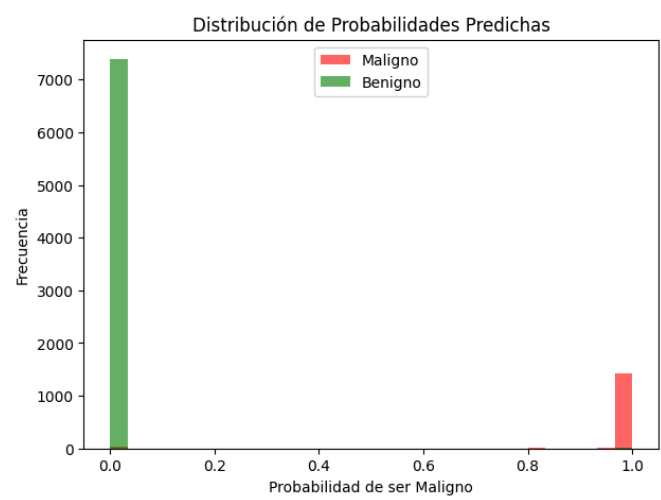


Figura 1. Distribución de Probabilidades

Donde podemos ver como las probabilidades estan distribuidas en los dos extremos.

### Evaluación comparativa en dos conjuntos de datos

Para verificar la capacidad de generalización del modelo final v4.3, se realizó una segunda evaluación utilizando imágenes de un conjunto de datos externo procedente de Kaggle. Este experimento permitió analizar cómo se comporta el modelo en un entorno diferente al utilizado durante el entrenamiento.

Las imágenes utilizadas en esta evaluación se distribuyen de la siguiente forma:

Conjunto de test de Database\_v2:

- 8.927 imágenes (7.420 benignas, 1.507 malignas)

Conjunto de Kaggle:

- 13.879 imágenes (7.289 benignas, 6.590 malignas)

Las métricas obtenidas en ambos conjuntos fueron las siguientes:

Conjunto de datos	F1 Score	Recall	Especificidad	Accuracy	AUC-ROC
DatabaseV2 (Test Set)	0,9755	0,9642	0,9974	0,9918	0,9965
Kaggle (Train + Test)	0,9618	0,9703	0,9572	0,9634	0,9884
Mediana	0,9687	0,9673	0,9773	0,9776	0,9925

Tabla 1. Métricas de los dos conjuntos

Los resultados muestran que el modelo mantiene un alto rendimiento en ambas bases de datos, aunque se observan algunas diferencias. En particular, la accuracy en Kaggle es ligeramente inferior a la obtenida en Database\_v2, lo que puede deberse a diferencias en la distribución y calidad de imágenes.

Sin embargo, el recall en Kaggle es mayor (97.03%), lo que sugiere que el modelo es más sensible a la detección de melanomas en este conjunto. Esto es un aspecto positivo, ya que confirma que el modelo prioriza la detección de casos malignos, reduciendo el riesgo de falsos negativos en entornos clínicos.

Para obtener una evaluación global del modelo, se calculó la mediana de los valores obtenidos en ambas bases de datos, proporcionando una referencia consolidada sobre su rendimiento general.

**F1 Score:** 0.9687

- Esta métrica combina precisión y sensibilidad, asegurando que el modelo tenga un equilibrio adecuado entre detección de melanomas y minimización de falsos positivos.

**Recall (Sensibilidad):** 0.9673

- Indica que el 96.73% de los melanomas fueron identificados correctamente.

**Especificidad:** 0.9773

- El 97.73% de las lesiones benignas fueron clasificadas correctamente.

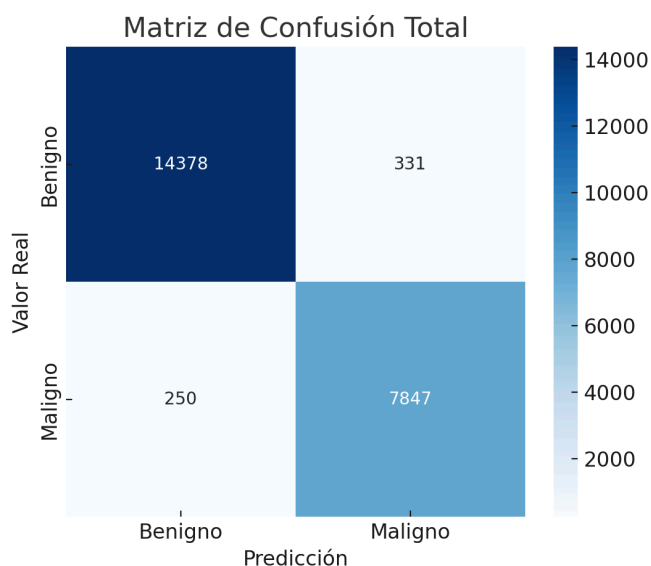
**Accuracy: 0.9776**

- La precisión global del modelo muestra que aproximadamente 98 de cada 100 imágenes fueron clasificadas correctamente.

**AUC-ROC: 0.9925**

- Este valor confirma una excelente capacidad de discriminación entre lesiones benignas y malignas en diferentes umbrales de decisión.

**Matriz de Confusión Resultante:**



**Figura. 2.** Ejemplo de pie de figura

El modelo clasificó correctamente 14,378 casos benignos (parte superior izquierda, azul oscuro), lo que indica que la mayoría de las muestras sin melanoma fueron identificadas correctamente. 331 casos benignos fueron clasificados erróneamente como malignos (parte superior derecha), lo que representa las falsas alarmas. Es preferible que el modelo tenga un ligero sesgo hacia la precaución antes que dejar pasar casos malignos sin detectar.

En cuanto a los melanomas, 7,847 casos fueron identificados correctamente (parte inferior derecha). Sin embargo, 250 melanomas fueron mal clasificados como benignos (parte inferior izquierda). Este es un aspecto crítico en aplicaciones médicas, ya que un falso negativo implica que un melanoma real no ha sido detectado.

Aunque la cantidad de falsos negativos es relativamente baja en comparación con el total de muestras, sigue siendo un aspecto que podría mejorarse.

## 9. Conclusiones

En este trabajo se ha desarrollado y evaluado un modelo de inteligencia artificial para la clasificación de lesiones cutáneas, explorando distintas arquitecturas y estrategias de optimización. Se han utilizado redes neuronales convolucionales (CNN) y modelos basados en Vision Transformers (ViT), aplicando técnicas avanzadas de preprocesamiento, data augmentation y optimización de hiperparámetros.

El modelo final desarrollado ha demostrado un rendimiento satisfactorio en términos de precisión, recall y especificidad, alcanzando valores elevados en ambas clases. La evaluación en múltiples conjuntos de datos ha permitido comprobar su capacidad de generalización, aunque se han identificado ciertos desafíos asociados a la variabilidad en la calidad y distribución de las imágenes.

Durante el desarrollo se ha constatado que el equilibrio entre precisión y recall es un aspecto crítico en aplicaciones médicas. La estrategia adoptada priorizó la detección de lesiones malignas, asegurando una baja tasa de falsos negativos a expensas de un ligero aumento en los falsos positivos. Esta decisión resulta adecuada en un contexto clínico, donde minimizar el riesgo de pasar por alto casos de melanoma es prioritario.

A pesar de los resultados obtenidos, existen aspectos susceptibles de mejora. El modelo se ha visto condicionado por limitaciones computacionales que han restringido la posibilidad de explorar configuraciones adicionales. Asimismo, se ha identificado la necesidad de ampliar la base de datos con imágenes más diversas y representativas para mejorar la capacidad de generalización sin depender exclusivamente de técnicas de aumento de datos.

## 10. Trabajo Futuro

El modelo desarrollado ha demostrado un rendimiento sólido en la clasificación de lesiones cutáneas, pero aún existen diversas áreas de mejora y optimización.

### Validación en entornos clínicos

Hasta el momento, el modelo no ha sido probado con imágenes recientes obtenidas en entornos clínicos reales. Un paso fundamental será establecer colaboración con hospitales y centros dermatológicos para validar su desempeño con imágenes de pacientes reales. Esto permitirá evaluar su efectividad en un contexto clínico y mejorar su capacidad de generalización con datos más diversos.

### Optimización de hiperparámetros y arquitectura del modelo

Si bien el modelo ha sido optimizado mediante técnicas avanzadas como Optuna, aún no se ha explorado la totalidad de combinaciones posibles de hiperparámetros. Una futura línea de trabajo consistiría en realizar una búsqueda más exhaustiva para encontrar una configuración aún más eficiente.

Además, se plantea la posibilidad de desarrollar una implementación personalizada de Vision Transformers (ViT) en lugar de utilizar modelos preentrenados. Esto permitiría realizar ajustes más específicos a las características de las imágenes dermatológicas. También se podrían explorar modificaciones en la arquitectura, como la congelación de capas o la incorporación de nuevas funciones de activación y pérdida.

### Infraestructura computacional y despliegue

Actualmente, el modelo no es completamente compatible con las versiones más recientes de TensorFlow para TPU en Google Colab, lo que representa una limitación para futuras mejoras. Será necesario actualizar la implementación para

garantizar su compatibilidad con versiones futuras y aprovechar mejor los recursos computacionales disponibles.

Por otro lado, se plantea la posibilidad de implementar el modelo en una infraestructura híbrida que combine procesamiento en la nube con ejecución local. Esto permitiría optimizar tanto el rendimiento del modelo como su disponibilidad en entornos clínicos, asegurando un equilibrio entre escalabilidad, seguridad y accesibilidad.

### Expansión de la base de datos

El conjunto de datos utilizado en este trabajo ha sido suficiente para entrenar el modelo con buenos resultados, pero su composición aún podría mejorarse. En particular, sería beneficioso incrementar la cantidad de imágenes de melanomas malignos para reducir la dependencia del data augmentation y mejorar la capacidad del modelo para detectar patrones reales en lugar de variaciones artificiales.

## 11. Agradecimientos

Este trabajo no habría sido posible sin el apoyo y la orientación de muchas personas, a quienes quiero expresar mi más sincero agradecimiento.

En primer lugar, quiero agradecer profundamente a mi profesor y tutor Paweł Gajewski, de la Universidad AGH en Cracovia, Polonia, por su dedicación, paciencia y constante apoyo a lo largo del desarrollo de este proyecto. Su experiencia y conocimiento han sido fundamentales en cada etapa del proceso, brindándome consejos y soluciones cuando más lo he necesitado.

También quiero agradecer a Xavier Masip, profesor en la Universitat Politècnica de Catalunya (EPSEVG), quien, aunque no participó activamente durante el desarrollo del proyecto, tuvo un papel clave en la etapa final de entrega de la memoria y preparación de la presentación. Su apoyo en esta fase ha sido de gran ayuda para cerrar el trabajo con éxito.

Extiendo mi gratitud a mi familia, por su apoyo incondicional, a mis amigos y compañeros, quienes han aportado ideas valiosas y me han acompañado durante todo el proceso. Gracias a su ánimo y aportaciones, este proyecto ha podido evolucionar y convertirse en una realidad.

A todos ellos, gracias por formar parte de este camino y contribuir a que este trabajo haya alcanzado su objetivo.

## Referencias

- [1] *Cáncer de piel*. National Cancer Institute. <https://www.cancer.gov/espanol/tipos/piel>
- [2] *Avance Redes Neuronales En Aplicaciones Médicas*. <https://consultorsalud.com/avance-de-las-redes-neuronales-en-aplicaciones-medicas/>
- [3] *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*. Harvard Dataverse. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T>
- [4] BCN20000. (n.d.). *BCN20000 Dermoscopic Lesions*. [https://figshare.com/articles/journal\\_contribution/BCN20000](https://figshare.com/articles/journal_contribution/BCN20000)
- [5] ISIC. (n.d.). *ISIC Melanoma Classification Data*. <https://www.kaggle.com/competitions/siim-isic-melanoma-classification/data?select=jpeg>
- [6] *Melanoma Skin Cancer Dataset of 10000 Images*. (n.d.). Kaggle. <https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images>
- [7] *Optuna: A hyperparameter optimization framework*. (n.d.). Optuna: A hyperparameter optimization framework — Optuna 4.2.1 documentation. <https://optuna.readthedocs.io/en/stable/>
- [8] *Qué buscar: ABCDE de melanoma*. (n.d.). American Academy of Dermatology. <https://www.aad.org/public/diseases/skin-cancer/el-abcde-del-melanoma>