TRABAJO FIN DE GRADO

INGENIERÍA INFORMÁTICA

Aplicación móvil para la prognosis y detección del cáncer de piel usando Deep Learning

Autor

Cristhian Moya Mota (alumno)

Directores

Diego Jesús García Gil

Julián Luengo Martín

Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación

—

Granada, 9 de junio de 2024

Aplicación móvil para la prognosis y detección del cáncer de piel usando Deep Learning

Autor
Cristhian Moya Mota

Directores
Nombre Apellido1 Apellido2 (tutor1)
Nombre Apellido1 Apellido2 (tutor2)

Título del Proyecto: Subtítulo del proyecto

Cristhian Moya Mota

Palabras clave: palabra_clave1, palabra_clave2, palabra_clave3, ...

Resumen

Poner aquí el resumen.

Project Title: Project Subtitle

First name, Family name (student)

Keywords: Keyword1, Keyword2, Keyword3, ...

Abstract

Write here the abstract in English.

Yo, **Nombre Apellido1 Apellido2**, alumno de la titulación TITULACIÓN de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI XXXXXXXXX, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Nombre Apellido1 Apellido2

Granada a X de mes de 201 .

D. **Nombre Apellido1 Apellido2 (tutor1)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

D. **Nombre Apellido1 Apellido2 (tutor2)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado ***Título del proyecto, Subtítulo del proyecto***, ha sido realizado bajo su supervisión por **Nombre Apellido1 Apellido2 (alumno)**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a X de mes de 201 .

Los directores:

Nombre Apellido1 Apellido2 (tutor1) **Nombre Apellido1 Apellido2 (tutor2)**

Agradecimientos

Poner aquí agradecimientos...

Índice general

Índice de figuras

Índice de tablas

1 Introducción

1.1. Motivación

El cáncer es una de las causas de muerte principales en el mundo. Su gran agresividad, así como su dificultad de diagnóstico, debido a su gran variedad de ubicaciones y manifestaciones, provoca que un alto porcentaje de casos no sean diagnosticados a tiempo correctamente. Tan solo en 2023, aproximadamente se registraron 20 millones de nuevos casos de cáncer a nivel mundial, y produciéndose algo menos de 10 millones de defunciones.

Estos registros provocan una gran inquietud en la población y entre los expertos de la materia; debido al aumento que se produce cada año, se espera que para el año 2050, el número de nuevos casos sea un 70 % mayor. Por desgracia, no existen formas de prevención claras para este tipo de enfermedad, ni un tratamiento efectivo que permita al paciente recuperarse fácilmente.

La única opción probada es la realización de pruebas rutinarias a colectivos de riesgo, para así acelerar la detección de posibles tumores, y aumentar la esperanza de supervivencia. Esto se ve reflejado en las cifras de los dos tipos de cáncer más frecuentes: el cáncer de mama, y el cáncer colorrectal. Si se detectan en fases iniciales, la correcta recuperación del colon podría aumentarse hasta el 90 %, mientras que en el cáncer de mama, podría reducirse su mortalidad entre un 25-31 %. Gracias a la existencia de pruebas rutinarias programadas por servicio de salud, se puede reducir la mortalidad.

El gran problema de estos tipos de cánceres son la escasa visibilidad y síntomas de los mismos; cuando muestran señales, es probable que la tasa de supervivencia sea mucho menor, sobre todo en el colon. Pero existe otro tipo de cáncer que sí se manifiesta de forma más visible y que puede alarmar al paciente de forma más temprana: el cáncer de piel.

Este tipo de tumores puede manifestarse en las diferentes capas de la dermis, y su origen se atribuye a la exposición prolongada a la luz solar sin hacer uso de protección. Debido a los daños que sufre la capa de ozono, y otros factores ambientales, la cantidad de rayos ultravioleta que llegan hasta la superficie ascendió desde que se tienen registros. Si bien la capa de ozono parece recuperarse, debemos ser cautos,

y tener cuidado de nuestra piel; los rayos ultravioleta pueden dañar células de la misma, y provocar alteraciones en su material genéctico. Son las que dan lugar al crecimiento incontrolado de células, son las que forman los tumores cancerígenos en la piel.

Se estima que en el mundo, los tumores de la piel representan un tercio de los casos de cáncer diagnosticados. Esta distribución sigue valores parecidos en España, y al igual que las cifras de otros tipos de tumores, los casos diagnosticados aumentan año tras año. Las muertes debidas a esta enfermedad son principalmente, por ser indentificadas en fases tardías de su evolución. Debido a que la piel es el órgano más grande del cuerpo humano, y que está en contacto con todos los capilares sanguineos y el sistema linfático, las células cancerosas se pueden extender por ellos hacia otros lugares del cuerpo.

Aunque este cáncer puede ser identificado de forma más sencilla por su portador, la escasa información acerca del tema, y la confusión con otras lesiones benignas de la piel como verrugas o lunares, provoca una disminución en las posibilidades de supervivencia. Por ello, el objetivo de este trabajo es aportar una nueva forma de diagnóstico que permita a los usuarios obtener una orientación acerca de qué posible lesión están experimentando en la piel, y sirvan como complemento del experto. O bien, ayudar a los expertos a tomar la decisión, acortando los tiempos de diagnóstico para aumentar las posibilidades de supervivencia. Esta tarea será realizada gracias al uso de uno de las herramientas en auge en la actualidad: la inteligencia articificial, y concretamente, el uso de DeepLearning para visión por computador.

Mediante una nueva arquitectura, el propósito es conseguir un buen modelo, capaz de segmentar las manchas de interés en la piel que estén recogidas en una fotografía. Dicha fotografía será capturada con el telefóno movil del usuario, retirando así la necesidad de disponer de dispositivos especializados. Y posteriormente, clasificar dichas manchas para ofrecer al usuario final una respuesta sólida acerca del posible tipo de lesión de piel que sufre.

1.2. El cáncer de piel

Prosiguiendo con el cáncer de piel, su diagnóstico si dificulta, sobre todo, por su amplia variedad de formas, tamaños, texturas y manifestaciones. Aunque su visibilidad pueda parecer evidente, (ya que es observable a nivel macroscópico) puede ser confundido fácilmente con lesiones benignas. Normalmente, suele dividirse entre dos tipos diferentes:

- **Melanomas de la piel**. Son la variante más peligrosa. Su origen se encuentra en los melanocitos, las células encargadas de dar el color bronceado a la piel. Éstas pueden comenzar a crecer sin control originando tumores, los cuales crecen y se diseminan rápidamente hacia otras regiones del organismo, provocando la metástasis, una extensión a nivel total del organismo. Es el más grave de los diagnósticos. Puede identificarse como una mancha oscura en la piel, formando tumores de color café oscuro. Sin embargo, debido a la gran variedad de reacciones, pueden darse de color rosado si dejan de producir melanina. Este aspecto dificulta su diagnóstico, por lo que el papel de las herramientas de visión por computador pueden ayudar a su identificación.
- **Cánceres no melanomas**. Este tipo de cánceres no se ubican en los melanocitos, y pueden ser tratados mediante otras ténicas menos agresivas debido a su rara probabilidad de expansión. Los más comunes, son los tumores de células basales y los de células escamosas:
  - Células basales. Componen la capa inferior de la piel, y son las células encargadas de sustituir aquellas que componen la capa más externa de la piel. Se encuentran , por tanto, en constante reproducción para cubrir aquellas que mueren en la superificie. Si experimentan alguna mutación, producen tumores de color similar al de piel del paciente, con la posibilidad de aparecer en colores como negro brillante en las pieles más oscuras.
  - Células escamosas. Son las células externas de la piel, con forma plana. Se regeneran constantemente gracias a las células basales, que producen estas células las cuales se aplanan a medida que ascienden hacia la capa externa. Es frecuente, de nuevo, en zonas expuestas al sol, sobre todo la cara. Normalmente, se encuentran bien localizados, y puede procederse a su extirpación. En casos en los que se haya extendido, se hace uso de radioterapia.

Aunque en base a su descripción parezcan distinguibles, son fácilmente confudidos por su variedad con otros tumores benignos de la piel, como:

- **Lunares(nevus)**: hiperpigmentación benigna en la piel.
- **Verrugas**: tumores benignos de piel, frecuente debido a virus como el del papiloma humano.

- **Lesiones vasculares**: varices, derrames, y otro tipo de problemas circulatorios.
- **Lipomas**: tumores de tacto blando, debido a su contenido en lípidos (grasa).
- **Queratosis seborreica**: son manchas cerosas, comúnmente desarrolladas en la espalda. De aspecto oscuro y gran relieve, no suponen ninguna amenaza más allá de posible incomodidad al roce o estética.

El uso de aprendizaje profundo para este fin resulta interesante como forma de mejora del diagnóstico ante casos malignos y benignos de gran similitud, los cuales pueden confunir y dificultar la labor incluso a expertos dermatólogos.

1.3. Dificultades del diagnóstico

La justificación de la realización de este proyecto se basa, sobre todo, en la dificultad de conocer la naturaleza del tumor del paciente. Habitualmente, se suele extraer una muestra del tejido afectado para proceder a su análisis en laboratorio. A esta técnica se le denomina biopsia.

Es un proceso efectivo, que consiste en el estudio bajo microscopio de las células extraídas, y el patrón que constituye el tejido. Su proceso más complejo es la extracción, ya que si ésta no se realiza correctamente, ciertas células cancerígenas pueden no aparecer en la muestra y realizarse un diagnóstico erróneo. Además, debido a la falta de especialistas en la materia que puedan realizar las incisiones, personal médico no cualificado en esta materia suele realizar su extracción. Se estipula que en lesiones inflamatarias, el correcto estudio de la patalogía se da en el 77 % de los casos si la muestra es recogida por un dermátologo, y un 41 % si la realiza un ayudante [ref].

Existen varias técnicas: mediante corte con tijera, mediante rasurado, extracción mediante bisturí, o en forma elíptica, persiguiendo la extracción total de la lesión. En el caso de las dos primeras opciones, solo está indicada si la lesión es superficial y no existe riesgo de que se trata de un melanoma por las complicaciones que esto conlleva de una posible diseminación y metástasis.

Aunque esta técnica suele ofrecer buenos resultados, debido a que la distinción entre un tumor de tipo melanoma, y un simple nevus puede ser complicada, sería de interés conocer una evaluación previa. Y es aquí donde podemos ver la utilidad del proyecto propuesto: se busca reforzar el diagnóstico del experto utilizándose el modelo entrenado con las imágenes de entrenamiento extraídas.

También hay que tener en cuenta otros factores que pueden perturbar el diagnóstico tradicional, como lo es la correcta manipulación de la muestra extraída, y su

coloración adecuada para mejorar el contraste del tejido y poder distinguir el patrón descrito por las células y su núcleo. Además, la herida dejada en la piel puede sufrir complicaciones como hemorragias o infecciones si no se tratan adecuadamente. Reducir por tanto este tipo de operaciones a las estrictamente necesarias gracias a un modelo de aprendizaje profundo es una opción atractiva.

1.4. Relevancia y objetivos del proyecto

Teniendo en cuenta los factores estudiados acerca de la enfermedad en los puntos anteriores, podemos afirmar que el uso de un modelo de aprendizaje profundo en smartphones permite:

- Minimizar los costes al hacer uso de un dispositivo esencial en nuestra vida diaria como los teléfonos móviles, aprovechando la posibilidad de que la manifestación de los tumores de piel son visibles a nivel macroscópico.
- Facilitar la toma de decisiones de los expertos dermatólogos en casos complicados, a modo de sistema de ayuda a la decisión.
- Realizar diagnósticos preliminares por parte del paciente en manchas o lesiones de procedencia desconocida para el mismo, aunque sigue siendo recomendable pedir cita a un experto.
- Defender la utilidad de los modelos de aprendizaje profundo en el estudio y evaluación de casos en el ámbito médico, siendo en este caso la identificación de la patalogía a nivel macroscópico, sin necesidad de realizar una biopsia con el posible riesgo que esto conlleva.
- Profundizar y mejorar el rendimiento de las redes convolucionales en dispositivos móviles, haciendo un uso responsable y optimizado de los recursos disponibles, teniendo en cuenta los modelos presentes en el estado del arte actual.

En este documento, estudiaremos el estado del arte actual, y analizaremos e implementaremos una aplicación para dispotivos móviles que sea capaz de realizar la segmentación de la mancha cutánea, y su posterior clasificación dentro de una lista de patologías posibles. Se busca informar al usuario del riesgo de su lesión, y de su posible diagnóstico final, el cual debe ser verificado por el experto. De esta forma, el software final puede contribuir a acelerar los diagnósticos de esta enfermedad, y favorecer la esperanza de vida de los pacientes en caso de que su lesión sea cancerosa.

2 Tendencias y Estado del arte

Antes de adentrarnos en el análisis del problema, debemos de tener en cuenta de que este problema es una temática en constante evolución, y por tanto, podemos encontrar diferentes conceptos y procedimientos seguidos en la literatura que pueden servirnos de inspiración para abordar el problema sin cometer los errores ya cometidos en el pasado, y ser capaces de encontrar un nuevo enfoque que nos ofrezca ventajas.

Generalmente, este problema ha sido abordado empleando hardware de computador de escritorio, por lo que la mayoría de modelos se centran en el aprovechamiento de los recursos hardware alojados en un servidor para realizar la clasificación y evaluación de las imágenes potencialmente cancerosas tomadas. Por tanto, nos adentraremos en sus conceptos, pero teniendo en cuenta que el proyecto propuesto hará uso de dispositivos móviles durante el tiempo de inferencia.

2.1. Aprendizaje profundo en dispositivos móviles

Con la creciente tendencia de la potencia de cálculo en los dispositivos actuales, prácticamente todos los aparatos electrónicos que nos rodean han crecido en cuanto a potencia y complejidad de cálculo. Los smartphones son precisamente uno de ellos, y nos acompañan cada día, por lo que es el dispositivo ideal para tareas de uso cotidiano y portabilidad.

Estos dispositivos, a diferencia de los computadores tradicionales, normalmente basado en la arquitectura x86 o AMD64, se basan en ARM, siguiendo como concepto de diseño ofrecer el máximo rendimiento posible dentro de unos consumos contenidos,mejorando el ahorro de energía y la pérdida de la misma mediante calor. El entrenamiento de modelos que encontramos habitualmente en la literatura, como ResNet, Inception o similares, es prácticamente inviable de forma nativa.

Sin embargo, en lo que respecta a la inferencia, éstos son capaces de ofrecer muy buenos resultados, gracias a la incorporación de hardware dedicado capaz de ofrecer estas características. Como prueba, podemos observar infinidad de aplicaciones que hace uso de ello, como Google Lens, que si bien se ayuda del uso de servidores de búsqueda especializados, es capaz de realizar procedimientos locales en los dis-

positivos de gama alta.

El objetivo del proyecto es aprovechar dicho vacío en la existencia de apliaciones de inferencia local para ofrecer una app que no necesite de conexión de red permanente para ofrecer resultados acerca de las manchas de piel identificadas.

Aprovechando el auge de los smartphones, grandes empresas, como Google y Meta, centran sus esfuerzos en la creación de arquitecturas basadas en redes convolucionales capaces en realizar detección de imágenes en tiempo real, para la clasificación de distintos objetos que podemos encontrar en nuestra vida cotidiana, y servir así como una herramienta de apoyo para diferentes necesidades. Sin embargo, esto es una tarea completa, ya que suelen carecer de complejas operaciones, o sacrificar en profundidad para lograr un rendimiento aceptable de unas decenas de milisegundos por inferencia.

A continuación, evaluaremos algunos de los modelos más conocidos y efectivos de propósito general, como SqueezeNet[8], MobileNet [6][16][5], ShuffleNet [22] y EfficientNet[19][4].

2.1.1. Squeeze-Net (2016)

Squeeze-Net[8] se centra en la reducción de complejidad de la arquitectura, sin pérdida de capacidad predictiva y evitando aplicar técnicas de compresión y cuantización[10] de modelos. Se autodefine como "una red al nivel de AlexNet[9] pero con una quincuagésima parte de los parámetros", haciendo alusión a disponer de una capacidad de cálculo similar a AlexNet, pero recortando en cuanto a número de parámetros necesarios.

Aunque el motivo de su creación no es de forma directa el uso de la arquitectura en dispositivos móviles, ha sido ampliamente utilizada en ellos al formar parte de la tendencia actual de reducción de coste computacional para reducir las necesidades de potencia de cálculo. De esta forma, se puede facilitar la implementación de redes convolucionales en sistemas empotrados con escasa capacidad de memoria y cómputo, haciendo uso de FPGAs [11].

Siguiendo este punto de vista, fue capaz de igualar e incluso superar levemente el rendimiento de AlexNet[9], empleando las siguientes técnicas:

- Uso de módulos **Fire**. Se trata de un nuevo tipo de estructura convolucional modular que puede ser apilado en capas al estilo de los módulos Inception [18] de Google. Consiste en una unidad modular ajustable en función de 3 parámetros: el número de convoluciones 1x1, y el número de filtros 1x1 y 3x3 de "expasión" a aplicar. El objetivo de añadir las convoluciones 1x1 es,

por un lado, la reducción de dimensionalidad del volumen a convolucionar, y por otro lado, la simplificación en número de parámetros. En arquitecturas de gran profundidad, como VGGNet o ResNet, quedó demostrado que este tipo de convoluciones permitían llegar más allá sin perder información relevante para el aprendizaje. 2.1.1

- Desplazamiento de los métodos de **reducción** de dimensionalidad hacia las capas más **profundas** de la arquitectura. En lugar de realizar pooling o aplicar stride a la hora de aplicar el filtro para reducir el volumen de salida en las primeras capas de la red, este tipo de transformaciones se reparten en capas más profundas para evitar que las capas cercanas al Head, de forma que se reduce la pérdida de características si retrasamos el subsampling del filtro.
- Eliminación de capas totalmente conectadas. Estas capas son, normalmente, las que mayor complejidad añaden al modelo por su gran cantidad de parámetros. Gracias al uso de Average Pooling en su última capa, podemos tener una red completamente independiente del tamaño de la entrada sin gran cantidad de parámetros ni necesidad de capas adicionales.

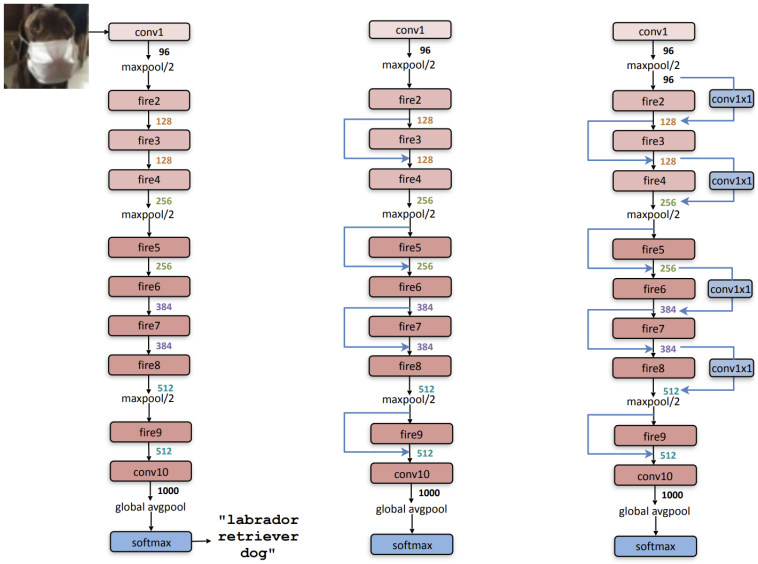


Figura 2.1: Arquitectura de SqueezeNet

Esta red ha sido empleada en multitud de aplicaciones: detección de objetos en tiempo real, clasificación semántica, y modelos preliminares de conducción autónoma. Entre todas sus aplicaciones, podríamos destacar su utilización en imagen médica, concretamente MRI (Resonancias magnéticas). Ha permitido facilitar el diagnóstico de ciertas enfermedades y lesiones cerebrales en un espacio de memoria y

recursos contenido.

Sin embargo, a pesar de las mejoras recibidas en sus versiones sucesivas, como los módulos Fire de doble nivel para reducir dimensionalidad, o la introducción de más reducciones mediante pooling, sacrifica resultados a nivel de accuracy respecto a la competencia, y no ha sido aplicada de forma firme y exitosa sobre imágenes de enfermedades cutáneas.

2.1.2. MobileNet

MobileNet es el fruto del proyecto de investigación de Google Research para la implementación de redes convolucionales en dispositivos móviles. El objetivo era encontrar un modelo eficiente que pueda ser incluso utilizado en tareas de segmentación en tiempo real, pero reduciendo el número de parámetros del red así como el número de operaciones de producto necesarias, para poder ejecutarlas de forma nativa en dispositivos móviles como smartphones y tablets.

Esta arquitectura consta de 3 versiones diferentes, siendo cada una más sofisticada que la anterior. Disponemos de MobileNet V1, MobileNet V2 y MobileNetV3.

MobileNet V1 (2017)

La versión original de la arquitectura convolucional MobileNet [6] fue publicada en 2017. En esta publicación, se busca reducir el número de operaciones realizadas para conseguir un menor impacto de las operaciones en punto flotante sobre el rendimiento.
El punto clave de esta arquitectura reside en las llamadas "pointwise convolutions", haciendo uso del concepto de separabilidad, ampliamente estudiado desde el año 2012 por la literatura.

Las nuevas convoluciones descomponibles se pueden separar en dos pasos bien delimitados: la convolución en profundidad y la convolución puntual.

- Las convoluciones en profundidad realizan el producto del filtro con el volumen de entrada capa a capa. Es decir, no se tiene en cuenta la dimensionalidad total de la imagen, sino que se realiza por cada nivel de profundidad el mismo producto. Esto reduce considerablemente el número de parámetros, ya que la dimensionalidad del problema es mucho menor.

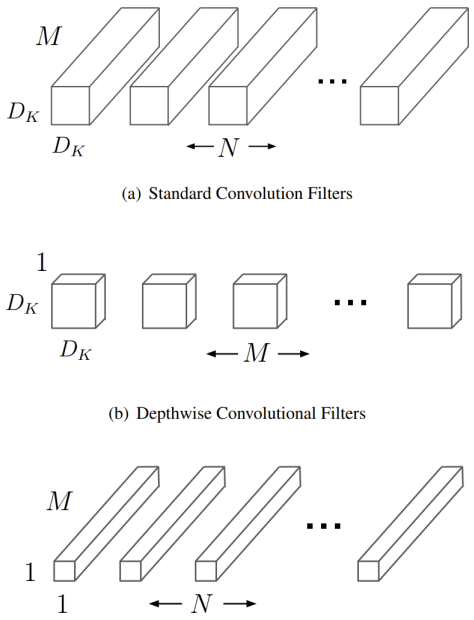
- La segunda fase es la convolución puntual, cuyo objetivo no es más que acumular el producto de todas las capas calculadas independientemente mediante una simple combinación lineal, la cual es de coste computacional muy bajo.

$$G_{k,l,m} = \sum_{i,j} K^{i,j,m} * F_{k+i-1,l+j-1,m}$$

Este producto es calculable eficientemente por las técnica de álgebra lineal GEMM, que permite aplicar propiedades de la suma y la multiplicación para el producto matricial de forma eficiente mediante Tensor cores.

En resumen, gracias a la separabilidad convolucional, se adquieren varias ventajas:

- El número de productos se reduce considerablemente. Como se puede verificar en [6] se traduce en una reducción de entre 8 y 9 veces el número de operaciones con respecto a las arquitectura tradicional de convolución
- Se reduce el espacio necesario en memoria.
- Se puede aprovechar el hardware específico.
- No se pierde precisión de cálculo gracias a que la separabilidad de convoluciones no afecta al resultado.



(a) Standard Convolution Filters

(b) Depthwise Convolutional Filters

(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Figura 2.2: Producto depthwise

Adicionalmente, también incorporaron dos hiperparámetros: α y ρ. Estos parámetros controlan la anchura y la resolución de entrada, respectivamente.

El hiperparámetro α hace referencia a la anchura de cada capa convolucional que compone la red, adquiriendo un valor de 1 cuando la arquitectura no se ve reducida, y gradualmente podrá ser reducida en el intervalo (0,1]. Así se conseguirán modelos más simples en anchura para dispositivos con menores recursos.

En cuanto a ρ, este se encuentra implícito en la resolución de la imagen de entrada. Por defecto, la red acepta imágenes de hasta 224x224, pero en función de dicho valor ρ, podremos reducir su resolución también dentro del intervalo (0,1].

Ambos parámetros sacrificarán bondad y ajuste en los resultados a favor de una mayor eficiencia.

MobileNet V2 (2018)

Dos años más tarde de la publicación de MobileNets, da a luz su versión V2 [16]. Esta conserva los hiperparámetros de la versión anterior, así como el producto punto a punto. Sin embargo, añade tres nuevas características, algunas de ellas no triviales y que requieren experimentación:

- Se introduce el concepto de "residuo invertido". Ésta mejora reside en la utilización de los bloques residuales, propuesto por la arquitectura de ResNet. Su objetivo es evitar la degradación del gradiente, y que se frene el aprendizaje en modelos de gran profundidad. Normalmente, estas conexiones se realizan entre capas de gran profundidad, siendo las capas intermedias bloques estrechos. Sin embargo, en MobileNet V2, se propone la composición inversa, de forma que sean los bloques intermedios entre los residuales aquellos que poseen una mayor anchura, y así reducir el número de parámetros sin perder expresividad en el modelo [12].

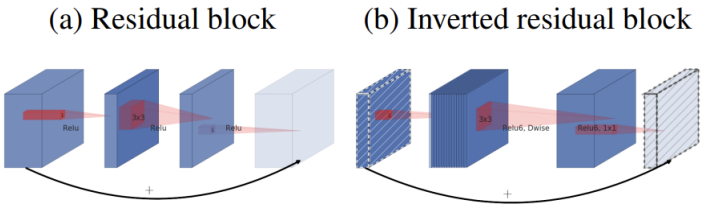


Figura 2.3: Bloque residual invertido

- En las capas donde el volumen de entrada es estrecho, al hacer uso de bloques residuales invertidos, la eliminación de la no linealidad aportada por ReLu favorece a la conservación de características y permite obtener mejores resultados de accuracy en tareas generales como clasificación en imagenet. Esto

se debe a que al realizar los saltos entre bloques "estrechos" perdemos rendimiento de la red, y simplemente con eliminar la última transformación no lineal del bloque, contrarrestamos este problema.

- ReLu6. Se mantiene una versión modificada de la original función de activación. En lugar de utilizar la tradicional función ReLu entre 0 y 1, se extiende este intervalo hasta 6, permitiendo mantener la precisión en caso de utilizar coma fija, ya que se aseguran 3 dígitos de parte entera, y el resto queda destinado a la mantisa, que se almacena de forma precisa.

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Figura 2.4: Arquitectura de MobileNetV2

MobileNet V3 (2019)

En su tercera versión[5], MobileNet incorpora métodos avanzados de diseños de redes basados en NetAdapt. Este algoritmo se basa en la transformación de modelos preentrenados para escritorio, y, en base a una serie de requisitos de potencia especificados, adapartar la arquitectura a una plataforma móvil perdiendo las mínimas capacidades posibles de la red original. El modelo de partida empleado fue ajustado con precisión para mejorar latencias y uso de memoria, aplicando los siguientes conceptos:

- Se añade la capa Squeeze-and-Excite, dentro de las conexiones residuales. Se trata de un mecanismo surgido en 2018 [7]. Este estudio afirma que existen filtros de imagen con mayor importancia para el cómputo global que otros, como, por ejemplo, los bordes. Por tanto, les aporta un mayor "peso" durante el entrenamiento a dichos filtros haciendo uso de una serie de parámetros adicionales. Éstos añaden una carga computacional muy pequeña, por lo que se trata de una técnica eficaz. Para obtener los parámetros de relevancia, se dispone de dos módulos: squeeze, y excite. El módulo squeeze se encarga de representar cada filtro mediante un valor numérico, obtenido por average

pooling de la imagen. Y por otro lado, el módulo excite se encarga de aprender los pesos que dar a cada uno de estos filtros o canales, haciendo uso de un MLP. El resultado final serán los pesos de cada canal en cuanto a su importancia, normalizados entre 0 y 1 por una función sigmoide.

- Se incluyeron nuevas capas al inicio y al final de la red de tipo residual invertidas.

Input	Operator	exp size	#out	SE	NL	s
$224^2 \times 3$	conv2d	-	16	-	HS	2
$112^2 \times 16$	bneck, 3x3	16	16	-	RE	1
$112^2 \times 16$	bneck, 3x3	64	24	-	RE	2
$56^2 \times 24$	bneck, 3x3	72	24	-	RE	1
$56^2 \times 24$	bneck, 5x5	72	40	✓	RE	2
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 3x3	240	80	-	HS	2
$14^2 \times 80$	bneck, 3x3	200	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	480	112	✓	HS	1
$14^2 \times 112$	bneck, 3x3	672	112	✓	HS	1
$14^2 \times 112$	bneck, 5x5	672	160	✓	HS	2
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	conv2d, 1x1	-	960	-	HS	1
$7^2 \times 960$	pool, 7x7	-	-	-	-	1
$1^2 \times 960$	conv2d 1x1, NBN	-	1280	-	HS	1
$1^2 \times 1280$	conv2d 1x1, NBN	-	k	-	-	1

Table 1. Specification for MobileNetV3-Large. SE denotes whether there is a Squeeze-And-Excite in that block. NL denotes the type of nonlinearity used. Here, HS denotes h-swish and RE denotes ReLU. NBN denotes no batch normalization. s denotes stride.

Input	Operator	exp size	#out	SE	NL	s
$224^2 \times 3$	conv2d, 3x3	-	16	-	HS	2
$112^2 \times 16$	bneck, 3x3	16	16	✓	RE	2
$56^2 \times 16$	bneck, 3x3	72	24	-	RE	2
$28^2 \times 24$	bneck, 3x3	88	24	-	RE	1
$28^2 \times 24$	bneck, 5x5	96	40	✓	HS	2
$14^2 \times 40$	bneck, 5x5	240	40	✓	HS	1
$14^2 \times 40$	bneck, 5x5	240	40	✓	HS	1
$14^2 \times 40$	bneck, 5x5	120	48	✓	HS	1
$14^2 \times 48$	bneck, 5x5	144	48	✓	HS	1
$14^2 \times 48$	bneck, 5x5	288	96	✓	HS	2
$7^2 \times 96$	bneck, 5x5	576	96	✓	HS	1
$7^2 \times 96$	bneck, 5x5	576	96	✓	HS	1
$7^2 \times 96$	conv2d, 1x1	-	576	✓	HS	1
$7^2 \times 576$	pool, 7x7	-	-	-	-	1
$1^2 \times 576$	conv2d 1x1, NBN	-	1024	-	HS	1
$1^2 \times 1024$	conv2d 1x1, NBN	-	k	-	-	1

Table 2. Specification for MobileNetV3-Small. See table 1 for notation.

Figura 2.5: Arquitectura de MobileNetV3

Debido a la gran complejidad adquirida por el modelo, los problemas de latencia y rendimiento en dispositivos de menor potencia, se opta por dividir la arquitectura en dos modelos parametrizables: MobileNet Small y Large. Mientras que la versión Large mejora los resultados de la versión 2 aumentando las prestaciones, el modelo Small otorga importancia sobre todo a la eficiencia y el uso de memoria, enfocado al hardware embebido o dispositivos de poca potencia.

Aplicaciones en dermatología y cáncer de piel

MobileNet, concretamente en su segunda versión, ha sido utilizado en la literatura para el diagnóstico de enfermedades de la piel. En [2], es utilizado para realizar la clasificación de 7 enfermedades cutáneas extraídas del Humans against Machine, HAM10000 [21], que podemos encontrar en ISIC archive [1], un repositorio web de acceso libre con enfermedades de la piel tanto beningnas como cancerosas.

También se utilizó más recientemente para su implementación en dispositivos de IOT, [17], donde se logra alcanzar el 99 % de accuracy en un pequeño conjunto extraido de ISIC, haciendo uso de la versión V3 junto a un algoritmo de Squeeze.

Dicho algortimo se encarga de localizar la ubicación de los pelos y otros posibles artefactos para preprocesar la imagen y lograr una fotografía resultante libre de interferencias.

Para ello, hace uso de un filtro black hat, que binariza la imagen y obtiene los píxeles objetivo de eliminar, que son sustituidos por los colores de los píxeles adyacentes, junto al uso del aumento de datos. Sin embargo, no queda especialmente claro la tasa de precisión del modelo para cada una de las enfermedades que se intentan diagnosticar.

2.1.3. Shuffle Net

Shuffle Net [22, 20] surge con el objetivo ofrecer un modelo capaz de ofrecer un buen modelo con la mínima pérdida de rendimiento frente a modelos profundos. Es capaz de superar en resultados a la primera versión de MobileNet, logrando un error de aproximadamente tres puntos menos que MobileNet V1. Su mejor rendimiento se debe sobre todo al uso de Channel Shuffle para las convoluciones grupales, y creando una arquitectura basada en módulos shuffle.

La convolución en grupo mediante mezcla de canales (Channel Shuffle) surge tras el estudio del funcionamiento de las convoluciones grupales en AlexNet[9] y ResNext. En ambos modelos, se uitilizan convoluciones grupales, donde cada canal de salida sólo se relaciona con los canales de entrada del que proviene. Esto podría debilitar la relación entre cada canal, y debilitar los resultados; para evitarlo, y no poner en riesgo el rendimiento con demasiadas convoluciones 1x1 para relacionarlos, se hace uso del mezclado de canales; es decir, es como si permitiésemos que cada grupo convolucional pudiera obtener información de otros grupos adyacentes, para así mejorar la relación entre ellos.

Para evitar un sistema complejo a la hora de representar dichas interconexiones, es como surge el channel shuffle 2.1.3: se mezclan los canales de forma que los grupos ya no quedan aislados con sus respectivas entradas y salidas.

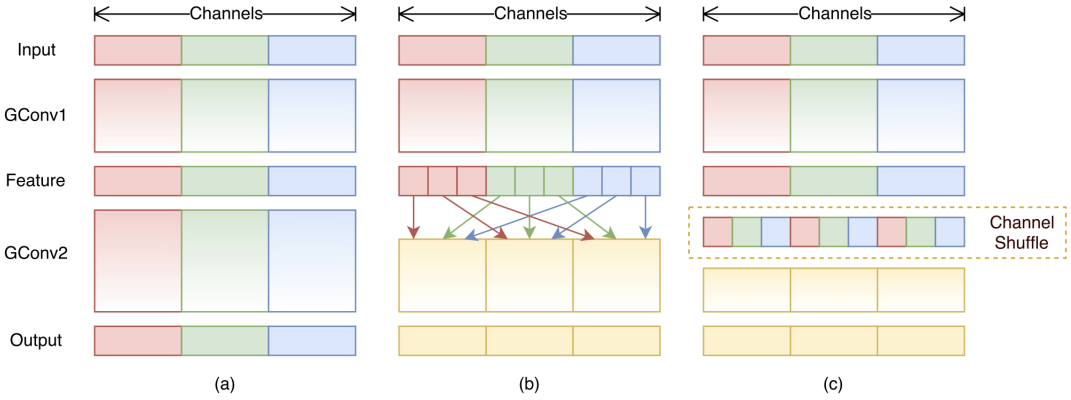


Figura 2.6: Channel Shuffle de ShuffleNet

Esta técnica se aplicará sobre la primera y última convolución 1x1 realizada sobre los bloques residuales de la red, que siguen una estructura parecida a la que adoptaría MobileNetV2 posteriormente. Mediante esta propuesta, podemos además aplicar una mayor capacidad de procesamiento en anchura añadiendo stride, y aplicando average pooling. El resultado, es conseguir modelos más anchos en procesamiento que no impacten negativamente en el rendimiento de los dispositivos con menor capacidad. En la figura **??** , podemos apreciar la arquitectura de la red, cuyos filtros pueden ser escalados mediante el parámetro s, aunque teniendo en cuenta una penalización en la complejidad, equivalente a s^2 sobre Shuffle Net base, equivalente a $s = 1$

Layer	Output size	KSize	Stride	Repeat	Output channels (g groups)				
					$g = 1$	$g = 2$	$g = 3$	$g = 4$	$g = 8$
Image	224×224				3	3	3	3	3
Conv1	112×112	3×3	2	1	24	24	24	24	24
MaxPool	56×56	3×3	2						
Stage2	28×28		2	1	144	200	240	272	384
	28×28		1	3	144	200	240	272	384
Stage3	14×14		2	1	288	400	480	544	768
	14×14		1	7	288	400	480	544	768
Stage4	7×7		2	1	576	800	960	1088	1536
	7×7		1	3	576	800	960	1088	1536
GlobalPool	1×1	7×7							
FC					1000	1000	1000	1000	1000
Complexity					143M	140M	137M	133M	137M

Figura 2.7: Arquitectura de ShuffleNet

Sus aplicaciones han sido variadas, pero en lo que respecta a la detección de lesiones cutáneas, la cercana salida de MobileNet V2 y su mejor rendimiento provocó

que ShuffleNet quedase relegada a un segundo plano, y no fuese muy utilizada para este fin. Podemos encontrar algunos trabajos [3] donde podemos observar una comparativa de este modelo frente a la completitud de los modelos del estado del arte de 2022, y podemos confirmar que MobileNetV2 es capaz de superar su rendimiento en la mayoría de pruebas, siendo estas comparaciones en cuanto a tiempo de entrenamiento, precisión, accuracy y tamaño del conjunto de entrenamiento. Sólo consigue superar a MobileNetV2 en tiempo de entrenamiento, donde es aproximadamente 900 más rápida, pero ofrece peores resultados en promedio.

2.1.4. EfficientNet

EfficientNet [2] es un conjunto de arquitecturas de redes creadas por el departamento de investigación de Google con el fin de conseguir una familia de modelos variada que fuese capaz de adaptarse fácilmente mediante parámetros a diferentes conjuntos de imágenes, y a requisitos de hardware más o menos limitados.

Parte de que una red convolucional sigue el siguiente esquema:

$$\mathcal{N} = \odot_{i=1\ldots s} F_i^{L_i}(X_{(H_i,W_i,L_i)})$$

Donde se denota que la capa F_i es repetida L_i veces la etapa i de la red, y la dimensionalidad de la capa queda representada con (W_i, L_i). Fijando F_i, efficient net intenta dar versatilidad a sus modelos variando las dimensiones restantes, Li, Ci, Hi, Wi mediante el uso de 3 constantes de escalabilidad:

- Profundidad, Depth (d): Aumentar la profundidad es la tendencia habitual presente en las redes convolucionales. Pero llegar a un equilibrio es crítico, ya que aumentar demasiado la profundidad sin modificar otros parámetros puede ocasionar pérdidades rendimiento por el desvanecimiento del gradiente a menor profundidad.

- Anchura, Width (w): Aumentar la anchura suele ser beneficioso para modelos de pocos recursos donde el aumento de profundidad supone un gran aumento de la carga computacional. Permite conseguir mayor cantidad de características de grado fino, pero si la red es demasiado poco profunda, el modelo careceré de características de alto grado que permitan aprender patrones generales.

- Resolución, (r): al emplear tamaños de entrada mayores, damos opciones a obtener una mayor cantidad de características de grado fino, pero un exceso de resolución puede provocar grandes tiempos de ejecución y puede ser contraproducente, al reducirse la ganancia con tamaños demasiados grandes.

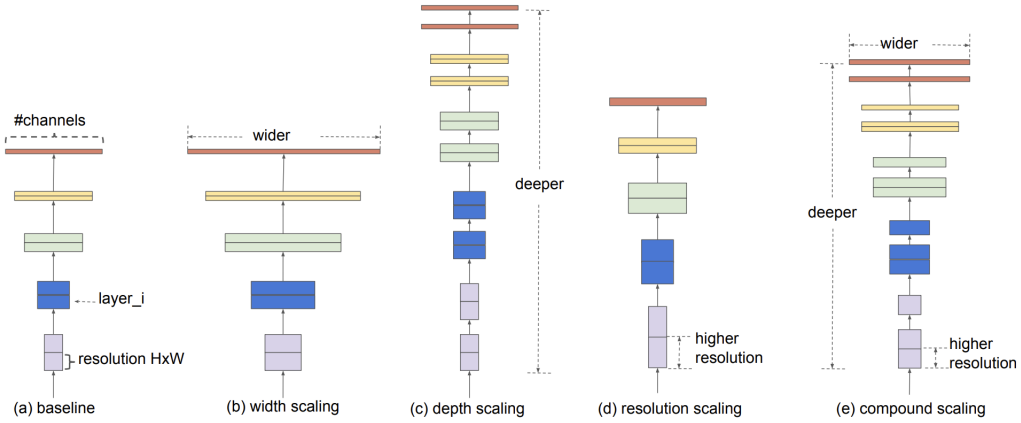


Figura 2.8: Parámetros de EfficientNet

Experimentalmente, estos parámetros pueden ser ajustados, y dan lugar a una serie de modelos distinto: los conocidos EfficientNetB0 - B7, denotando el valor numérico la profundidad y complejidad del modelo, siendo esta mayor a mayor valor del índice. Cada una de ellas fue ajustada utilizando como requisito la potencia medida en TFLOPS para su ejecución, y puede ser posteriormente ajustada con el resto de parámetros libres no fijados a las características del conjunto de entrada.

En smartphones de alta gama, los modelos B0 a B4 pueden ser ejecutados con un rendimiento aceptable para aquellas aplicaciones que no requieran un alto tiempo de respuesta, pero si necesitan dar al usuario una respuesta aceptable. Para modelos de mayor complejidad computacional, se usan las variantes lite, que estudiaremos más adelante.

Aplicaciones en dermatología y cáncer de piel

En el problema que nos concierne, esta arquitectura ha conseguido grandes resultados en el dataset del ISIC abierto al público como competición en la plataforma Kaggle, habiendo sido utilizado como parte de un ensemble de modelos, o bien como modelo único entrenado en el top 3 de ganadores de la competición. En el caso de la segunda mejor solución clasificada [14], se menciona la utilización de EfficientNet-B6, con tamaño de entrada de 512x512, y un tamaño de batch de 64, obteniendo 0.9485 de accuracy como resultado final a la hora de emplear los datasets ISIC 2019 y 2020
En la primera solución, es usada en conjunto a Resnet50, y una red especializada en los metadatos de la imagen, y todos los modelos juntos someten su resultado a votación [13].

Ambos resultados han sido evaluados con computadores de alta gama, haciendo uso de múltiples tarjetas gráficas para el entreanmiento y la inferencia. Este proceso

es demasiado pesado para un dispositivo móvil, por lo que de cara a este trabajo, se buscarán alternativas capaces de ahorrar en espacio y potencia como concepto de cuantización.

2.2. Cuantización de modelos

Las arquitecturas y soluciones propuestas por la literatura que han sido analizadas en los apartados anteriores proponen métodos que ofrecen buenos resultados. Sin embargo, existe un problema: todas ellas han sido entrenadas con un computador cuya capacidad de cálculo supera incluso las características de un ordenador doméstico promedio, como en [14], donde se emplean 4 GPU Nvidia Quadro RTX 6000 24GB. Aunque su utilización engloba sobre todo el proceso de entrenamiento, la inferencia de estos modelos también sigue siendo extremadamente costosa para un dispositivo móvil, y este proceso también es realizado a través del computador.

Esto resume a que el teléfono simplemente adquiere el papel de cliente dentro de una arquitectura cliente-servidor, donde el dispositivo host de todo el procesamiento de la imagen y de su clasificación es un ordenador de gran potencia de cálculo, y el teléfono movil únicamente debe compartir con este la imagen que desea examinar. Pero esto supone una gran desventaja: en ausencia de conexión de red, o interrupción del servicio por parte del servidor, el usuario no sería capaz de emplear la aplicación para el diagnóstico. La dificultad e interés de este proyecto se ve reforzado por este argumento, y resulta ahora de interés la posibilidad de realizar la inferencia en el teléfono, a pesar de que el entrenamiento sea realizado en un ordenador.

Existe una técnica que nos permitirá obtener un modelo optimizado a partir de uno entrenado de forma tradicional: la cuantización.

2.2.1. Características

La cuantización de modelos se centra en la simplificación y optimización de un modelo preentrenado por un computador, reduciendo algunas de sus características buscando un impacto mínimo sobre los resultados obtenidos, pero reduciendo de forma considerable el tiempo de inferencia para dispositivos de baja potencia. Aunque existen mecanismos de poda, donde la arquitectura del modelo se ve simplificada de forma directa, el método que evaluaremos será la cuantización de modelos basada en la reducción de la precisión numérica, es decir, una disminución en la profundidad en bits de la variable flotante.

Este concepto ya fue brevemente mencionado durante el desglose de características de MobileNet V3 [5]. En la arquitectura AMD64, el almacenado de variables en coma flotante emplea una precisión de 32 bits. Por tanto, al realizar operaciones de cualquier tipo, las 32 posiciones del nuevo número han de ser actualizadas al valor

resultado. El tiempo empleado en realizar dicha operación en cada dígito de este número puede parecer despreciable, pero, cuando el conteo de operaciones alcanza los miles de millones, supone una diferencia significativa en el rendimiento.

En dispositivos de baja potencia, esta precisión suele verse reducida a 8 bits, de forma que reducimos la longitud del máximo número almacenable en 4 veces menos, y reducimos también así el tiempo por operación.

Este mecanismo es empleado tanto por los frameworks de trabajo habituales para aprendizaje mediante redes convolucionales, como Pytorch y TensorFlow, y por los propios fabricantes de teléfonos móviles de forma nativa. Es el caso de Qualcomm, conocido por su gama de procesadores Snapdragon. Esta empresa realizó un estudio del impacto de la cuantización en los modelos [10] usando flotantes de 8 bits del precisión. Demuestran que el uso de números flotantes de 8bits en lugar de enteros con exponenciación para desplazar el punto ofrece un mayor rendimiento. Las consideraciones a la hora de defender la utilidad de la cuantización son las siguientes:

- La operación más costosa realizada durante la inferencia es el producto de matrices. Para simplificar la complejidad de la operación, normalmente se suelen emplear matrices de números enteros reescalados a un nuevo intervalo de valores. Esto es, se aplica una transformación de rangos, donde una matriz $\mathbb{R}^{mxn}$ se cuantiza a una matriz $X^{(int)}$ asociado a un valor de reescalado s: $X^{(int)} = clip([\frac{X}{s}], x_{min}, x_{max})$, donde el producto pasa a ser el valor redondeado más cercano al número tras aplicar la transformación de rango mediante escala, y la operación de clip asegura que el número es representable dentro de los valores establecidos para el extremo. Esto proporciona una simplificación contra la habitual notación IEEE-754 de 32-bit, FP32, donde se usa un bit de signo, 23 de mantisa, y 8 bits de exponente.

- La utilización de FP8 puede ser beneficiosa al ser capaz de encontrar un término medio entre la precisión y variabilidad de rango numérica del resultado, como se puede ver en [10]. Los resultados empíricos muestran que tras un ajuste entre el número de bits de mantisa y exponente, se logra un equilibrio adecuado usando valores cercanos a 3 bits de mantisa y 4 de exponente, y en caso de post training quantization, 5 de mantisa y 2 de exponente, en caso de que el aprendizaje se realice de forma lenta.

- Para mejorar estos resultados, se puede usar la técnica Quantization Aware Training (QAT), o entrenamiento preparado para cuantización, donde el objetivo es facilitar el paso a FP8 desde el proceso de entrenamiento teniendo en cuenta los rangos de valores adoptados durante el entrenamiento, como por ejemplo, controlando el diferencial mínimo que puede tomar el algoritmo de optimización. Esta operación muestra mejor resultado sobre valores enteros, pero también es capaz de mejorar los resultados obtenidos por PF8.

En los frameworks Pytorch y TensorFlow, podemos encontar estas optimizaciones en sus funcionalidades relacionadas con la optimización de modelos en post entrenamiento, con las dependencias de Pytorch Mobile y TensorFlow lite, respectivamente. Ambas aplican cuantización numérica de los tensores, y en caso de tratarse de modelos tradicionales de la literatura, existen arquitecturas simplificadas de forma profunda, con redes como EfficientNet o ResNet, que cuentan con variantes lite.

2.2.2. Modelos cuantizados

El framwork de TensorFlow lite contiene dos de los modelos más utilizados en la literatura en sus variantes lite: EfficientNet Lite y ResNetV2. Ambos han recibido un tratamiento de simplificación, que pasa por el uso de INT8 como modelo de representación numérica y la simplificación de la arquitectura, retirando algunas de las capas más resentidas en el proceso de optimización. En el caso de EfficientNet Lite, podemos encontrar un descripción detallada sobre aquellos cambios realizados para la mejora de rendimiento [15]:

- Se hace uso de cuantización post entrenamiento mediante el algoritmo de cuantización empleado por Tensorflow lite. Esta cuantización se hace de forma dinámica, dependiendo del hardware en el que se ejecuta el modelo. Puede realizarse una cuantización de rango dinámico, donde el factor de escalado puede aplicarse de forma distinta dependiendo de la capa en la que se realice la operación; cuantización de enteros, para los dispositivos menos potentes; y por último, cuantización de FP16, de forma que se logra un término medio entre ambas soluciones, y se puede emplear en dispositivos con GPU de potencia considerable. Por defecto, es la opción de rango variable la utilizada.

- Se eliminan algunos módulos Squeeze and Excite presentes en capas intermedias, debido a su coste e imprecisión para dispositivos que no soportan gran precisión numérica.

- Las funciones de activación se reemplazan con RELU6, al igual que se realizó en MobileNetV3 [5]

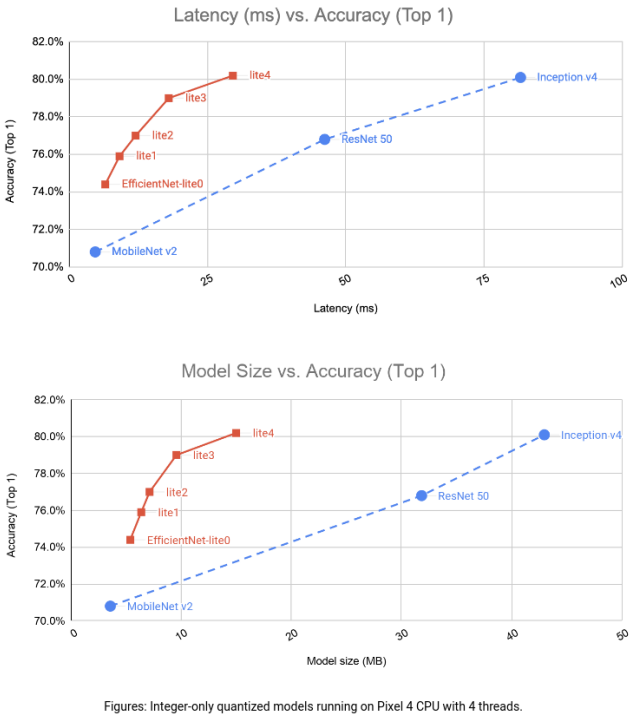


Figura 2.9: Optimización de EfficientNet Lite

Este procedimiento variable, con múltiples posibles combinaciones, se realiza debido a la gran variedad del mercado actual. Existe una gran diversidad de dispositivos: conviven dispositivos de bajo consumo con escasa potencia gráfica, así como terminales especializados con unidades de TPU para acelerar el cálculo neuronal. Y, de esta forma, con optimizaciones modulares, podemos priorizar la eficiencia en dispositivos con poca capacidad de cálculo , y aumentar la precisión en aquellos que pueden ejecutarlos. En el dispositivo de ejemplo, un Google Pixel 4, se consigue un tiempo medio de inferencia de 30 ms, una mejora de casi el 60 % sobre ResNet50.

2.2.3. Conclusión de los modelos cuantizados

A la vista de los resultados, los modelos cuantizados pueden suponer una ventaja; obtenemos resultados cercanos a los modelos complejos del estado del arte, con una pequeña penalización ganada en forma de rendimiento, y que la ejecución ofrezca tiempos de respuesta razonables. Además, este tipo de transformaciones no se han empleado para el reconocimiento y detección de enfermedades de la piel, y abren una nueva linea de investigación en la que obtener resultados.

Como puntos en contra, debemos tener en cuenta que la penalización obtenida al entrenar este tipo de modelos no es siempre la misma; dependiendo del problema a clasificar, y del modelo seleccionado, la penalización de la cuantización puede

ser o no más marcada; para ello, es necesario obtener empiricamente resultados que nos permitan conocer el tipo adecuado para nuestro problema. Este conflicto para encontrar el equilibrio ya fue detectado por los desarrolladores de Google en EfficientNet Lite [4, 15], donde inicialmente, sus pruebas en ImageNet arrojaron resultados pésimos al obtener un 46 % de accuracy en clasificación TOP1 a diferencia del 75.1 % obtenido con FP32. Pero, tras ajustar el rango de amplitud de los enteros utilizados para cuantización, se consiguió una ganancia de 1.85x sobre el modelo original con tan solo un 0.7 % de pérdida en los resultados.

	float32	int8	Improvement
model size	17.7MB	5.17MB	3.42x
accuracy (top1)	75.1%	74.4%	-0.7 percent point
latency (CPU)	12ms	6.5ms	1.85x

* Benchmarked on Pixel 4 CPU with 4 threads

Figura 2.10: Ganancia de la cuantización en EfficientNet

Como aspecto negativo, aunque la varianza de los resultados pueda verse paliada mediante la selección de la cuantización más adecuada, aún existen riesgos de resultados inesperados. Con este mismo modelo, por ejemplo, se han alcanzado resultados inesperados, como que los modelos de menor tamaño (B0,B1,B2) poseen un mejor desempeño que B3 y B4 en datasets de propósito general, como podemos observar en el artículo de Agarwal [**?**], pero esto puede deberse por la necesidad de una mayor cantidad de datos para realizar el entrenamiento.

Dado a los buenos resultados en general, la cuantización será la senda seguida por este TFG a la hora de optimizar los modelos de escritorio en post entrenamiento.

2.3. Recursos gráficos disponibles

La obtención de datos es un proceso fundamental en la resolución de problemas de Machine Learning. Este tipo de problemas requieren un gran número de imágenes que aporten variedad, y permitan construir un modelo general que se capaz de adaptarse a cambios de iluminación, diferentes puntos de vista y composiciones.

Es clave, por tanto, disponer de diferentes tipos de lesiones, tanto benignas como malignas, así como diferentes tonos de piel. La inexistencia de un tipo de piel en el conjunto de entrenamiento, o la inexistencia de un tipo de lesión podrían provocar resultados sesgados indeseados durante la predicción de la imagen tomada.

Podemos encontrar en la red varios datasets de acceso público que permiten su utilización de forma abierta con fines académicos. Dada a la gran cantidad de publicaciones disponibles, resulta complejo averiguar si los datos a los cuales hace referencia se encuentran disponibles públicamente, si son de acceso restringido, o

bien, ya no se encuentran disponibles debido a cambios en su política o la falta de mantenimiento.

Debido a que el estudio de la evolución y el diagnóstico del cáncer de piel de forma temprana es un tema en auge, existen gran cantidad de publicaciones especializadas únicamente en el análisis de los conjuntos de datos públicamente accesibles, como es el caso de la lista propuesta por M. Goyal et Al. [16], o el reciente estudio realizado por Sana Nazari y Rafael García (2023)[30]. Basados en su modelo de estudio y las referencias recomendadas por sus artículos, se ha elaborado el siguiente plan de búsqueda para saber qué tipo de datos utilizar y cuáles descartar:

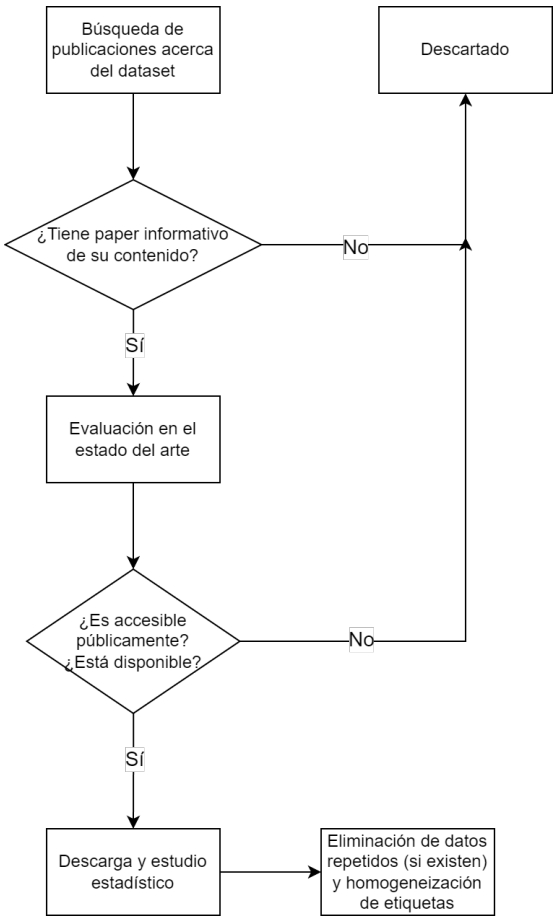


Figura 2.11: Proceso de búsqueda seguido

Siguiendo dicho procedimiento, se han encontrado 6 datasets diferentes, cuyo origen son instituciones públicas que han cedido datos con fines académicos y de investigación.

Un factor determinante para la elección de estos conjuntos es la diversidad. Es

indispensable, para este proyecto, encontrar datos lo suficientemente variados como para distinguir lesiones cancerígenas y no cancerígenas, y disponer de diferentes tonos de piel para entrenar. Aunque los tonos de piel más oscuras sufren lesiones de tipo cancerígeno en menor proporción gracias a su protección natural, también pueden sufrir este tipo de patologías, y es clave ser capaces de detectarlas en cualquier posible paciente.

ISIC Data

El repositorio ISIC (International Skin Imaging Collaboration [1]), contiene imágenes demoscópicas de lesiones principalmente cancerosas. Existe una gran cantidad de publicaciones acerca de este conjunto de datos, debido a su utilización anual durante los años 2016-2020 para la realización de un reto virtual de Machine Learning descrito en [1]. El objetivo, consiste en identificar los melanomas frente a lesiones no cancerosas (ISIC Challenge-2020[2]) o bien, identificar diferentes subtipos de lesiones cancerosas frente a lesiones benignas (ISIC Challenge 2019, [3]).

En el estado del arte, destacan las soluciones que hacen uso de métodos de DeepLearning, como el planteado por Ian Pan [4], finalista de la competición para el año 2020. El ganador de la competición del año 2020, cuyo análisis podemos encontrar en [5], realiza por su parte un enfoque híbrido entre el uso de DeepLearning para la clasificación de los datos de tipo imagen, y clasificación mediante Machine Learning para los metadatos, y así obtener un resultado más preciso.

Debido a que cada uno de estos subconjuntos de datos anuales podían ser pequeños, normalmente se recurría a la reutilización de los conjuntos anteriores para enriquecer el conjunto de entrenamiento. Este método es bastante recomendable, ya que, a mayor conjunto de entrenamiento, más cercanos se encontrarán los parámetros de interés del problema general a solucionar. Sin embargo, hay que realizar dicha fusión con especial cuidado, ya que existen volúmenes de datos considerables que se repitieron en las competiciones de cada año para aumentar el tamaño del dataset, y si se realizase una simple concatenación de los datos, estaríamos desperdiciando esfuerzo computacional en clasificar imágenes redundantes (comportamiento para nada deseable al trabajar sobre entornos móviles de menor potencia.).

Un análisis extenso de los datos asociados a cada Challenge podemos encontrarlo en [6]. En él, se recogen otros modelos del estado del arte utilizables para este fin, así como una forma de tratar los datos duplicados. Los datos pueden ser separados en los siguientes subconjuntos:

Challenge Dataset Year	Train	Test	Total	Tipo de problema
ISIC 2016	900	379	1279	Clas. binaria
ISIC 2017	2000	600	2600	Clas. Multiclase
ISIC 2018	10015	1512	11527	Clas. Multiclase
ISIC 2019	25331	8238	33569	Clas. Multiclase
ISIC 2020	33126	10982	44108	Clas. Binaria

- ISIC 2016 [7]: Es el dataset de menor tamaño de todos los propuestos. Hace distinción únicamente de los casos malignos y benignos. Contiene imágenes dermoscópicas anotadas con información acerca de la localización de la mancha, y la edad del paciente. Contiene información adicional para la segmentación de la mancha pigmentada de interés (máscaras).
- ISIC 2017 [8] Es un conjunto de mayor tamaño al anterior, y hace alusión a 4 clases diferentes: melanomas, nevus, y seborrheic keratosis. Contiene también información acerca de la edad del paciente, y otros metadatos de interés. La escasa cantidad de datos provoca que normalmente en la literatura este dataset se utilice también como clasificación binaria entre nevus y keratosis, u otros enfoques similares.
- ISIC 2018 [9]. Este dataset contiene un número de imágenes considerables, siendo un total de 10015 imágenes para entrenamiento, y 1512 para test. En este caso, se realiza subclasificación de tipos, a través de las clases melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma y lesiones vasculares. Es de especial interés destacar que este dataset proviene, a su vez, de HAM10000 (Human against machine, [11]) y MSK Dataset [12]. El challenge original comprendía, de nuevo, la clasificación de los diferentes tipos realizando previamente una discriminación de la mancha en cuestión mediante segmentación. Existen gran cantidad de publicaciones que tratan este conjunto de datos, como [13], donde se emplea este dataset para demostrar mejores resultados al emplear transformaciones polares de la imagen y aumentar la invarianza.
- ISIC 2019 [2]. Se trata del mayor conjunto de datos para clasificación multiclase propuesto por ISIC [1]. Se trata del mismo dataset que el año 2018 [9], con la adición de BCN_20000 Dataset [14], cuyos datos provienen del Hospital Clínic de Barcelona [14]. Las clases a clasificar se amplían hasta 9, encontrando subtipos de melanomas en el conjunto.
- ISIC 2020 [3]. El último dataset propuesto públicamente, contiene únicamente datos binarios acordes a melanomas y no malignos.

Todos estos datos pueden ser acoplados entre sí para dar un dataset global de ISIC [6], donde obtendríamos las siguientes clases:

Clase	2017	2018	2019	2020
Melanoma	374	1113	4522	584
Atypical melanocytic proliferation	-	-	-	1
Cafe-au-lait macule	-	-	-	1
Lentigo NOS	-	-	-	44
Lichenoid keratosis	-	-	-	37
Nevus	-	-	-	5193
Seborrheic keratosis	254	-	-	135
Solar lentigo	-	-	-	7
Melanocytic nevus	-	6705	12.875	-
Basal cell carcinoma	-	514	3323	-
Actinic keratosis	-	327	867	-
Benign keratosis	-	1099	2624	-
Dermatofibroma	-	115	239	-
Vascular lesion	-	142	253	-
Squamous cell carcinoma	-	-	628	-
Other / Unknown	1372	-	-	27.124
Total	2000	10.015	25.331	33.126

Sin embargo, sería necesario tener en cuenta la eliminación de imágenes repetidas, debido a que durante cada edición de ISIC, un número considerable de imágenes han sido incluidos en varios años. Este procedimiento engloba:

1. Eliminar las imágenes idénticas por hash. Todas las imágenes de ISIC están numeradas de forma única para facilitar la identificación de cada una de ellas. Si unimos todos los datatesets, y tomamos las repeticiones, podemos remover:

	2016	2017	2018	2019	2020
Train	291	1283	0	0	0
Test	95	594	0	0	0

Tabla 2.1: Número de imágenes duplicadas recogidas por [6]

2. Eliminación del ISIC 2018. Como éste se encuentra contenido en la composición para el año 2019, puede prescindirse totalmente de él a favor de la versión de 2019.

3. Eliminación de imágenes "downsampled" del conjunto. En los años 2019 y 2020, se añadieron imágenes de challenges anteriores con una reducción en resolución. Para ahorar en espacio y tiempo de cómputo, pueden eliminarse las imágenes reducidas para quedarnos con una única copia de mayor calidad

de la lesión, y luego realizarles manualmente un reescalado en caso de que sea necesario.

Atendiendo de nuevo a los resultados propuestos por [6], obtenemos el siguiente conjunto:

Year	Task No.	Images Removed	Images Remaining
2016	3	826	74
2017	3	801	1199
2018	3	10,015	0
2019	1	2235	23,096
2020	-	433	32,693
Total	-	14,310	57,0621

Tabla 2.2: Tabla de imágenes únicas extraída de [6]. En este caso, el autor descarta el uso del dataset de 2016 por su baja aportación

Obtendríamos un total de 57000 imágenes, los cuales podrían clasificarse, con sus respectivas clases extraídas de los metadatos. Componen, en resumen, un conjunto de datos robusto que puede formar parte del dataset de entrenamiento de este trabajo.

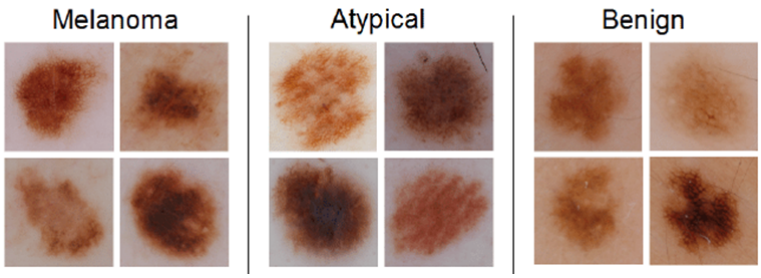


Figura 2.12: Ejemplo de imágenes de ISIC 2017 [23]

ASAN Dataset

ASAN (Seung Seog Han 2018)[15][18][19] es un conjunto de datos de origen surcoreano compuesto por lesiones malignas y benignas de la piel. Nos permite obtener un mayor grado de variedad de las imágenes, ya que el repositorio ISIC se centra sobre todo en lesiones de piel de población europea.

Tipo de lesión	Número de ejemplares
Actinic keratoses and intraepithelial carcinoma (AKIEC)	651
Basal Cell Carcinoma (BCC)	1082
Dermatofibroma (DF)	1247
Hereditary angioedema (HAO)	2715
Intraepithelial Carcinoma (IC)	918
Lentigo (LEN)	1193
Melanoma (ML)	599
Nevus (NV)	2706
Pyogenic Granuloma (PG)	375
Squamous Cell Carcinoma (SCC)	1231
Seborrhoeic Keratosis (SK)	1423
Wart	2985
Total	**17125**

Tabla 2.3: Distribución de clases de ASAN dataset

Tal y como se describe en [17] (M Goyal 2019), este dataset tiene 12 tipos de enfermedades, sumando un total de 17125 imágenes clínicas. Estas imágenes están compuestas en su mayoría por imágenes en miniatura, pero existe un repositorio con imágenes de mayor tamaño, donde sería necesario realizar tareas de segmentación- Sin embargo, dichas imágenes son de acceso restringido, y se requieren permisos especiales del hospital para acceder a ellos. Por ese motivo, tendremos únicamente en cuenta loas 17125 miniaturas.

Adicionalmente, podemos encontrar también imágenes proporcionadas por Hallym, un dataset complementario de 125 imágenes pertenecientes a lesiones de tipo melanoma cancerosas.

Las clases más destacadas de este dataset en su conjunto son la presencia de lesiones benignas de la piel, detalle que no encontramos en ISIC, y que permiten así contrastar información de la piel con lesiones benignas con la piel cancerosa. Podemos encontrar 4 clases benignas: lentigos (manchas solares fruto del envejecimiento y la exposición prolongada al sol), nevus (lunares comunes), verrugas y granulomas benignos.

Figura 2.13: Ejemplo de lunares beningnos en ASAN (Nevus)

Los resultados han sido confirmados por expertos dermatólogos y los resultados verificados en su mayoría mediante biopsia, por lo que las etiquetas asociadas a cada lesión están completamente verificadas.

El formato de las imágenes es una disposición matricial de las miniaturas, donde cada fichero que contiene las subimágenes representa en su conjunto una clase. Por desgracia, no se aporta otro tipo de información adicional más allá de la etiqueta por motivos de privacidad.

Dermnetz

Podemos encontrar extraer este dataset de un atlas online de enfermedades cutáneas recogidas de pacientes alrededor de todo el mundo. Contiene tanto lesiones benignas como malignas, existiendo además manchas y lesiones vinculadas a enfermedades infecciosas y hongos.

Existen gran cantidad de herramientas para realizar esta extracción de datos, como la que podemos encontrar en [21]. En total, se pueden obtener hasta 23000 imágenes, existiendo un total de 23 clases no balanceadas. Podemos encontrar lesiones de tipo alérgico, así como acné, dermatitis severa o celulitis.

Carece de metadatos asociados, ya que dicha información es de carácter reservado por su mantenedor. El dataset no se encuentra listo para usar de forma inmediata, ya que desde 2019, las imágenes deben ser extraídas de la propia web, pues dispone un índice donde se pueden acceder a las enfermedades de interés. El fichero contenedor del dataset fue retirado en 2019 del libre acceso, junto a sus metadatos. Es necesario

solicitar su acceso y aportar una cantidad económica.

PH2

El conjunto de datos PH2 [22] es un conjunto de 200 imágenes obtenidas gracias al hospital Pedro Hispano de Portugal. Está compuesto por imágenes de alta resolución que contienen 3 posibles casos de lesiones:

- Lunar común (Common Nevus), 80 ejemplares
- Lunar atípico (Atypical Nevus), 80 ejemplares
- Melanomas, 40 ejemplares.

Además de las 200 imágenes, podemos encontrar metadatos asociados a cada una de ellas, como el color, su extensión, textura, forma del borde, localización, entre otros. Su acceso es libre para fines académicos desde su página oficial [22], que contiene las imágenes en formato jpg, y varios ficheros .csv con la información de la imagen y su clasificación.

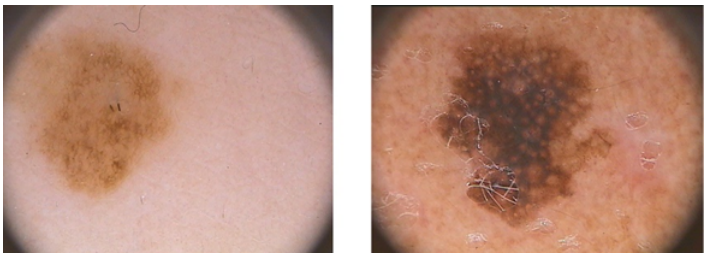


Figura 2.14: Nevus maligno y benigno en PH2

PAD-UFES 20

PAD-UFES-20 [25] se trata de un conjunto de datos recopilado de diferentes poblaciones, que contiene diagnósticos para 1.641 lesiones cutáneas únicas recopiladas, comprendiendo un total de 2.298 imágenes.

Entre sus clases, podemos encontrar tres enfermedades y tres cánceres de piel. Todos estos datos han sido recogidos y verificados mediante biopsia en un 100 % de los casos cancerosos, por lo que su diagnóstico está totalmente verificado.

Podemos encontrar, además del diagnóstico, metadatos acerca de:

- ID de paciente
- ID de lesión,

- ID de imagen
- Si la lesión benigna fue o no probada por biopsia.
- Información del paciente: fumador o no, localización de la lesión, edad, exposición a químicos, historial cancerígeno, etc.

Los datos han sido recogidos mediante teléfonos móviles en formato PNG, siendo las imágenes validadas por el Hospital Pathological Anatomy Unit of the University Hospital Cassiano Antȳnio Moraes (HUCAM) de la Federal University of Espírito Santo (Brasil). En su publicación original [25], podemos encontrar un resumen de su contenido de forma más específica:

Diagnostico	**Ejemplares**	**% biopsied**
Actinic Keratosis (ACK)	730	24.4 %
Basal Cell Carcinoma of skin (BCC)	845	100 %
Malignant Melanoma (MEL)	52	100 %
Melanocytic Nevus of Skin (NEV)	244	24.6 %
Squamous Cell Carcinoma (SCC)	192	100 %
Total	**2298**	**58.4 %**

Tabla 2.4: Tabla de casos diagnosticados en PAD-UFES20

Donde podemos apreciar que todos los casos de enfermedades cancerígenas han sido probados mediante biopsia, y el cáncer de célula basal se trata del tipo de enfermedad más frecuente.

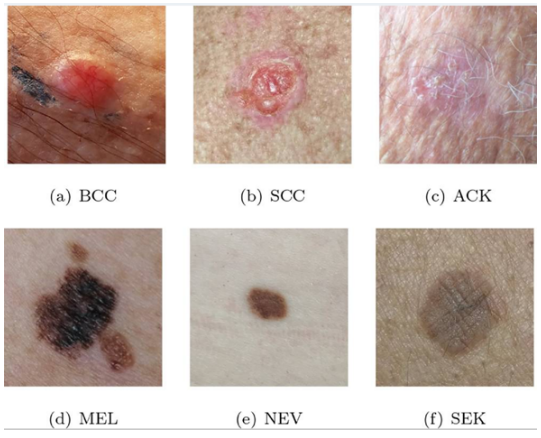


Figura 2.15: Batch de ejemplo de PAD-UFES 20 [25]

Severance

Se trata de un conjunto de imágenes de lesiones cutáneas [24] recopiladas de pacientes de Corea del Sur. Recibe dicho nombre a que los datos recopilados cuentan con la colaboración del Hospital Severance, en el mismo país.

En su variante A, que es la única disponible públicamente, podemos encontrar el diagnóstico y otra información asociada sobre 10426 imágenes, cuya valoración se encuentra entre las 38 posibles clases que contiene este conjunto de datos.

Seleccionando las 6 clases más comunes contenidas en este dataset, encontraremos que comprenden aproximadamente el 75 % del conjunto. Está compuesto por actinickeratosis (22.5 %), angiofibromas (14.4 %), angiokeratomas(13.8 %), cáncer de tipo basal cell (8.1 %), Becker nevus (7.5 %), bluenevus (6.2 %), y la enfermedad de Bowen (carcinomas)(6.1 %).

El interés en este dataset se debe a que algunas de estas clases mayoritarias, como los nevus azules y de Becker, son condiciones benignas que suelen ser retirados únicamente con fines estéticos, permitiendo complementar con el resto de los diagnósticos negativos. Este tipo de lunares son los más complejos de diagnosticar, debido a sus colores similares a un melanoma, y suelen requerir una biopsia, por lo que su diagnóstico suele alargarse.

Las imágenes se encuentran en formato matriz, por lo que es necesario proceder a su separación previo a su utilización con fines de Deep Learning:

Figura 2.16: Imágenes de ejemplo provenientes del dataset Severance

Otros datasets

Existen otros datasets ampliamente referenciados que son de acceso público. Sin embargo, en los últimos años, éstos han sido retirados y han quedado inaccesibles. Es el caso de DermQuest, un atlas virtual que contenía lesiones cutáneas y otras patologías. Ese dataset fue contenido posteriormente por Derm101, pero ambas versiones

fueron retiradas para su descarga. Alternativamente, podemos encontrar algunas de sus imágenes en los datasets SD-198 y SD-260 [26][29], pero únicamente permanece en activo el primero de ellos, bajo solicitud. En total, SD-198 contiene más de 6500 imágenes, mientras que SD-260 alcanzaba las 20000 imágenes.

En el estado del arte actual, podemos encontrar otros datasets ampliamente utilizados, como el caso de DermIS [27], un atlas online de patologías de la piel. También existen publicaciones reciente sobre nuevos conjuntos de datos utilizados de uso restringido, que permiten observar que la tendencia de investigación de este campo sigue en alza; es el caso del estudio propuesto por Papadakis et Al (2021) [28], que recoge datos sobre pacientes con melanoma de grado 3 para estudiar su evolución durante un período de 3 años, para estimar su crecimiento y potencial grosor del tumor.

Debido a las restricciones de acceso, ninguno de estos datasets será empleado como parte del entrenamiento del modelo diseñado para este estudio.

Conjunto resultado

Una vez examinados todos los conjuntos mencionados anteriormente, podemos llevar a cabo la unión de todos los datos en un único subconjunto. Esto nos permitirá conseguir un dataset completo y variado con diferentes tipos de piel y diferentes lesiones que nos permitirán identificar multitud de tipos de patologías, siendo posible ajustar el grado de granularidad en función de la agrupación o no de posibles subclases.

Inicialmente, el conjunto de datos construido contendrá todos los subtipos de lesiones cutáneas vistos, pero dispondrán de una segunda etiqueta que indicará si se trata de un caso canceroso o no, atendiendo a su subclase que lo etiqueta. Si agrupamos por lesiones benignas, cancerosas, y potencialmente cancerosas, obtenemos:

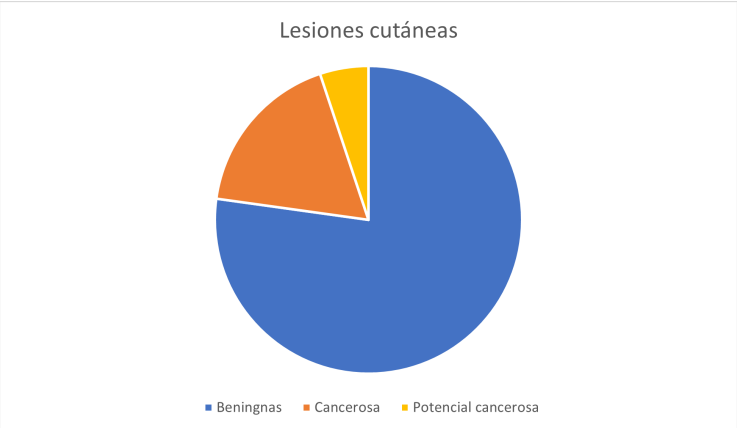


Figura 2.17: Distribución de clases

Se puede observar cómo la mayoría de imágenes disponibles engloban problemas de piel no cancerosos, mientras que el segundo tipo más común de lesión si es la cancerosa. Si atendemos a clasificar las subclases de cada tipo de patología, encontramos 52 posibles etiquetas.

Bibliografía

[1] The international skin imaging collaboration. https://www.isic-archive.com, 2023. [Online; accessed 20-September-2023].

[2] Saket S. Chaturvedi, Kajol Gupta, and Prakash S. Prasad. *Skin Lesion Analyser: An Efficient Seven-Way Multi-class Skin Cancer Classification Using MobileNet*, page 165–176. Springer Singapore, May 2020.

[3] Mohammad Fraiwan and Esraa Faouri. On the automatic detection and classification of skin cancer using deep transfer learning. *Sensors*, 22:4963, 06 2022.

[4] Google. Efficientnetlite (mediapipe. https://console.cloud.google.com/vertex-ai/publishers/google/model-garden/efficientnetlite?hl=es-419&pli=1, 2024. [Online; accessed 18-April-2024].

[5] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3, 2019.

[6] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.

[7] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks, 2019.

[8] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size, 2016.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[10] Andrey Kuzmin, Mart Van Baalen, Yuwei Ren, Markus Nagel, Jorn Peters, and Tijmen Blankevoort. Fp8 quantization: The power of the exponent, 2024.

[11] Samyukta Lanka. Megha Arora. Accelerating squeezenet on fpga. https://lankas.github.io/15-618Project/, 2024. [Online; accessed 21-April-2024].

[12] Paul-Louis Pröve. Mobilenetv2: Inverted residuals and linear bottlenecks. https://towardsdatascience.com/mobilenetv2-inverted-residuals-and-linear-bottlenecks-8a4362f4ffd5, 2018. [Online; accessed 21-April-2024].

[13] Paul-Louis Pröve. Siim-isic melanoma classification, 1st solution. https://www.kaggle.com/competitions/siim-isic-melanoma-classification/discussion/175412, 2020. [Online; accessed 8-June-2024].

[14] Paul-Louis Pröve. Siim-isic melanoma classification, 2nd solution. https://www.kaggle.com/competitions/siim-isic-melanoma-classification/discussion/175324, 2020. [Online; accessed 8-June-2024].

[15] Google Renjie Liu. Higher accuracy on vision models with efficientnet-lite. https://blog.tensorflow.org/2020/03/higher-accuracy-on-vision-models-with-efficientnet-lite.html, 2020. [Online; accessed 18-April-2024].

[16] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2019.

[17] Rupali Kiran et al. Shinde. Squeeze-mnet: Precise skin cancer detection model for low computing iot devices using transfer learning, 2022.

[18] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.

[19] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020.

[20] Sik-Ho Tsang. Review: Shufflenet v1 — light weight model. https://towardsdatascience.com/review-shufflenet-v1-light-weight-model-image-classification-5b253dfe982f, 2019. [Online; accessed 8-June-2024].

[21] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 08 2018.

[22] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices, 2017.