

Universidad de La Habana
Facultad de Matemática y Computación



Aprendizaje automático orientado a la clasificación de cáncer de piel. Un enfoque basado en EfficientNetB1

Autor:

Deborah Famadas Rodríguez

Tutores:

**Dr. Reinaldo Rodríguez Ramos, Universidad de la
Habana**

Dr. Yudivian Almeida, Universidad de la Habana

Trabajo de Diploma
presentado en opción al título de
Licenciado en Ciencia de la Computación

Fecha

github.com/deborahfam/Thesis

A todas esas personitas hermosas que me ayudaron a llegar aquí. Este mérito es
suyo.

Agradecimientos

Agradecimientos

Opinión del tutor

Opiniones de los tutores

Resumen

El Machine Learning (Aprendizaje Automático) es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones [1]. Las Redes Neuronales Convolucionales (CNN), una especialización de esta disciplina, conocidas por su eficacia en el procesamiento de imágenes, son ideales para detectar características comunes en imágenes dermatológicas.

Este trabajo de diploma propone un algoritmo de machine learning, específicamente, una red neuronal, que permite extraer, a partir de un conjunto de imágenes de cáncer de piel, características para luego realizar una clasificación.

Abstract

Machine Learning is a discipline in the field of Artificial Intelligence that, through algorithms, gives computers the ability to identify patterns in massive data and make predictions. Convolutional Neural Networks (CNN), a specialization of this discipline, known for their efficiency in image processing, are ideal for detecting common features in dermatological images.

This diploma work proposes a machine learning algorithm, specifically, a neural network, which allows extracting, from a set of skin cancer images, features to then perform a classification.

Índice general

| | |
|---|-----------|
| Introducción | 1 |
| 1. Estado del Arte | 5 |
| 1.1. Teledermatología | 5 |
| 1.2. Técnicas tempranas de clasificación de imágenes | 6 |
| 1.2.1. Clasificación y reconocimiento de patrones en imágenes médicas | 7 |
| 1.3. Introducción y desarrollo de machine learning en el campo médico . . | 8 |
| 1.3.1. Análisis de imágenes médicas | 8 |
| 1.4. Datasets de cáncer de piel | 9 |
| 1.5. Redes neuronales convolucionales en la medicina | 9 |
| 1.6. Retos y limitaciones actuales | 12 |
| 1.7. Conclusión del estado del arte | 13 |
| 2. Propuesta de solución | 14 |
| 2.1. Desarrollo del modelo y estrategias de optimización | 14 |
| 2.1.1. Selección del modelo y transferencia de aprendizaje | 14 |
| 2.1.2. Preprocesamiento de datos y balanceo de clases | 16 |
| 2.1.3. Arquitectura del Modelo y Regularización | 16 |
| 2.1.4. Ajuste dinámico del learning rate | 16 |
| 2.1.5. Diversidad en propuestas de soluciones | 17 |
| 3. Detalles de implementación y experimentos | 18 |
| 3.1. Enfoque de Implementación | 18 |
| 3.1.1. Herramientas y Tecnologías | 18 |
| 3.2. Preparación y carga de datos | 19 |
| 3.2.1. Dataset | 19 |
| 3.2.2. Transformación de datos | 20 |
| 3.2.3. Modelación y división del conjunto de datos | 20 |
| 3.2.4. Generadores de datos y preprocesamiento | 20 |
| 3.2.5. Aumento y división de datos | 20 |
| 3.3. Diseño y entrenamiento del modelo | 21 |

| | | |
|-----------|---|-----------|
| 3.3.1. | Capas adicionales y regularización | 21 |
| 3.3.2. | Optimización | 23 |
| 3.3.3. | Ajuste dinámico de la tasa de aprendizaje | 23 |
| 3.4. | Experimentos | 23 |
| 3.4.1. | Experimento 1: Evaluación de la eficiencia de la división asimétrica de datos en la clasificación de imágenes de cáncer de piel | 23 |
| 3.4.2. | Experimento 2: Análisis de la estratificación de datos en la clasificación de imágenes de cáncer de piel | 25 |
| 4. | Resultados | 27 |
| 4.1. | Experimento 1: Evaluación de la eficiencia de la división asimétrica de datos en la clasificación de imágenes de cáncer de piel | 27 |
| 4.2. | Experimento 2: Análisis de la estratificación de datos en la clasificación de imágenes de cáncer de piel | 32 |
| 4.3. | Comparación entre los Experimentos | 35 |
| 4.3.1. | Comparación de Ventajas y Desventajas | 36 |
| 4.4. | Consideraciones Finales | 37 |
| | Conclusiones | 38 |
| | Recomendaciones | 39 |
| | Bibliografía | 41 |

Índice de figuras

| | |
|--|----|
| 2.1. Estadísticas del rendimiento de los modelos de EfficientNet | 15 |
| 4.1. Estadísticas básicas del modelo EfficientNetB1. | 29 |
| 4.2. Estadísticas de aprendizaje a lo largo del proceso de entrenamiento. . | 29 |
| 4.3. Estadísticas de eficacia del modelo al estimar los resultados en el conjunto de pruebas | 30 |
| 4.4. Gráfico de errores por clase en el conjunto de pruebas | 31 |
| 4.5. Estadísticas básicas de algunas iteraciones del modelo EfficientNetB1. | 32 |
| 4.6. Estadísticas de aprendizaje a lo largo del proceso de entrenamiento. . | 33 |
| 4.7. Estadísticas de aprendizaje a lo largo del proceso de entrenamiento. . | 33 |
| 4.8. Estadísticas de eficacia del modelo al estimar los resultados en el conjunto de pruebas | 34 |
| 4.9. Gráfico de errores por clase en el conjunto de pruebas | 35 |

Ejemplos de código

Introducción

Desde su surgimiento en 2006, el aprendizaje profundo se ha establecido como una importante subdisciplina dentro del aprendizaje automático, especialmente en áreas relacionadas con la percepción visual humana. Esta metodología procesa datos a través de múltiples capas que incluyen estructuras complejas y transformaciones no lineales [29]. Actualmente ha logrado avances significativos en áreas como la visión artificial, el reconocimiento de voz, el procesamiento del lenguaje natural, el reconocimiento de audio y la bioinformática [19]. Desde 2013, el aprendizaje profundo se ha reconocido como uno de los diez avances tecnológicos más significativos, dadas sus amplias aplicaciones potenciales en el análisis de datos [9].

El enfoque del aprendizaje profundo abstrae los datos en distintos niveles, lo que permite su aplicación en tareas complejas como la detección de objetos y la clasificación. Su capacidad para reemplazar la extracción manual de características por algoritmos eficientes de aprendizaje, ya sea no supervisada o semi-supervisada, ha revolucionado múltiples áreas [42]. Esta revolución incluye el campo de la atención médica, donde la gestión y análisis de un volumen abrumador de datos médicos es un desafío crucial.

En el ámbito de la atención médica, especialmente en la dermatología, el aprendizaje profundo ha mostrado un potencial extraordinario. La dermatología, que se enfoca en el estudio y tratamiento de enfermedades de la piel, se enfrenta al desafío del cáncer de piel, el tipo de cáncer más frecuente a nivel mundial. La detección precoz de este cáncer es vital [24], y aquí es donde el aprendizaje profundo, con su habilidad para analizar y clasificar imágenes médicas con precisión, juega un rol transformador. La integración de estas tecnologías en la práctica dermatológica no solo mejora la precisión diagnóstica, sino que también promete revolucionar el tratamiento y manejo de diversas afecciones cutáneas [20]."

Motivación

La motivación detrás del uso de algoritmos de *machine learning* (ML) para el diagnóstico del cáncer de piel es significativa y valiosa. A diferencia de otros tipos de

cáncer, el cáncer de piel se forma en la superficie de la piel y suele ser visible. Esto plantea una oportunidad única para la detección temprana y el tratamiento, lo cual es esencial ya que la mayoría de los casos de cáncer de piel son tratables si se detectan a tiempo [24].

Estos algoritmos pueden identificar patrones complejos con una precisión y consistencia mayor que los métodos de diagnóstico humano, reduciendo así la posibilidad de diagnósticos incorrectos debido a la interpretación subjetiva y variable de los expertos [16]. Además, el ML puede procesar grandes volúmenes de datos rápidamente y su capacidad para aprender y adaptarse con el tiempo significa que la detección y clasificación del cáncer de piel puede mejorar continuamente [16].

Aunque ya existen algoritmos eficientes para la clasificación de melanomas, una forma de cáncer de piel, el desarrollo de un modelo capaz de clasificar varios tipos de cáncer de piel y generalizar entre ellos es un objetivo crucial. Esto ampliaría el alcance de las imágenes dermatológicas procesables, mejorando potencialmente la precisión en la detección y el tratamiento de distintas formas de cáncer de piel."

Antecedentes

El desarrollo de las Redes Neuronales Convolucionales (CNN) ha sido fundamental para la identificación de características en imágenes médicas, una base sobre la cual se construye la motivación actual para aplicar ML en la dermatología. Su uso en medicina ha demostrado ser eficaz para capturar patrones específicos en datos de imágenes con alta precisión ([4]). Por ejemplo, Brinker et al. [8] en 2018, analizan 13 artículos sobre la aplicación de CNN en la clasificación de lesiones cutáneas, resaltando su alto rendimiento y la posibilidad de reutilizar CNNs pre-entrenadas. En 2020, Ameri et al. [5] hacen un avance significativo al implementar una CNN profunda para procesar imágenes dermatoscópicas directamente, lo que mejora la eficacia del proceso de clasificación. Más recientemente, en 2022, Shetty et al. [40] logran una precisión del 95,18% en la clasificación de lesiones cutáneas utilizando CNN, demostrando su superioridad sobre otros algoritmos de ML.

En otras investigaciones enfocadas al diagnóstico mejorado de lesiones pigmentadas, Tajerian et al. (2023) [43] presentan un enfoque metodológico para el diagnóstico de lesiones cutáneas pigmentadas utilizando CNN, logrando una alta puntuación F1 de 0,93. En 2021 Adegun et al. [3] recoge un conjunto de estudios enfocados en el desarrollo de algoritmos con CNN para la detección de cáncer de piel. Estos incluyen enfoques de segmentación y clasificación, arquitecturas de auto-encoder-decoder, y la implementación de redes pre-entrenadas como AlexNet y VGG16.

Las universidades han desempeñado un papel crucial en la investigación y el desarrollo de tecnologías avanzadas en el campo de la medicina, especialmente en la detección y tratamiento del cáncer. Estos centros académicos no solo proporcionan

una base sólida para la investigación teórica, sino que también fomentan la innovación práctica mediante el uso de tecnologías emergentes como el aprendizaje automático y la inteligencia artificial. En particular, nuestra universidad ha contribuido significativamente a este campo.

Las universidades han desempeñado un papel crucial en la investigación y el desarrollo de tecnologías avanzadas en el campo de la medicina, especialmente en la detección y tratamiento del cáncer. Estos centros académicos no solo proporcionan una base sólida para la investigación teórica, sino que también fomentan la innovación práctica mediante el uso de tecnologías emergentes como el aprendizaje automático y la inteligencia artificial. En particular, nuestra universidad ha contribuido significativamente a este campo.

Un claro ejemplo de esta contribución es la tesis de Darien Viera Barredo titulada *Autómata celular estocástico en redes complejas para el estudio de la invasión, migración y metástasis del cancer* [49] proporciona un marco detallado sobre cómo se aborda el estudio del cáncer desde una perspectiva matemática y computacional avanzada. El modelo propuesto utiliza autómatas celulares estocásticos para simular el crecimiento avascular y vascular del tumor. En el se aborda la complejidad del ciclo vital tumoral, destacando la importancia de su comprensión tanto para la investigación del cáncer como para la salud pública. Tradicionalmente, la modelación matemática y computacional se ha centrado en las etapas tempranas del desarrollo tumoral, donde la mortalidad es baja. Sin embargo, este estudio se enfoca en las fases avanzadas, que son críticas para la vida del paciente.

Complementando esta línea de investigación, la tesis reciente de Claudia Olavarrieta Martínez [33] propone un *ensemble* de redes neuronales para clasificar imágenes dermatoscópicas en cuatro categorías: melanoma, carcinoma basocelular, carcinoma espinocelular y otros utilizando la técnica de transferencia de conocimientos en redes como VGG16, ResNet50 y EfficientNet B0.

Problemática

Es entonces notorio que las investigaciones mencionadas utilizan técnicas de machine learning para llevar a cabo el proceso de clasificación. Esto tiene sentido dado que la problemática central en la detección del cáncer de piel radica en la necesidad de mejorar la precisión y rapidez del diagnóstico. Tradicionalmente, esta tarea recae en dermatólogos y el diagnóstico de melanoma, que depende de la evaluación clínica y los hallazgos clásicos en la biopsia de la lesión. Pero la inspección visual puede no ser suficiente para diferenciar lesiones benignas de tumores malignos, y aunque la biopsia de piel es el estándar de oro, es un procedimiento invasivo con limitaciones. Además, la experiencia, el costo y la disponibilidad son desafíos para el uso generalizado de herramientas no invasivas en el diagnóstico clínico [16].

Objetivos

Este trabajo propone como objetivo fundamental el diseño y validación de un modelo predictivo basado en *deep learning* para el diagnóstico del cáncer de piel mediante la clasificación de imágenes dermatológicas. El modelo diseñado y desarrollado se enfoca en clasificar imágenes, priorizando tanto la precisión de los resultados como la capacidad de generalización del modelo. Esto se concibe así con la idea de desarrollar trabajos posteriores que admitan otros datos de entrada.

Entre los objetivos específicos del proyecto se encuentran:

1. Estudiar el estado del arte sobre las técnicas empleadas en el diagnóstico de imágenes dermatológicas y su efectividad.
2. Crear un modelo de *deep learning* que dado un conjunto de imágenes de cáncer de piel clasifique el tipo al que pertenecen.
3. Decidir mediante un algoritmo de machine learning la mejor distribución de datos para entrenamiento del modelo e implementar mejoras potenciales al modelo e hiperparámetros para mejor precisión del mismo.
4. Implementar técnicas de validación para evaluar la precisión del modelo.

Contribuciones

La metodología propuesta podría contribuir al desarrollo de un sistema de clasificación de imágenes de cáncer de piel más preciso, para luego ser utilizado en la práctica clínica en el diagnóstico de cáncer de piel. De esta forma podrían extenderse los algoritmos de clasificación existentes para el análisis de estos datos.

Estructura de la tesis

El contenido de la tesis se organiza de la siguiente forma. En el capítulo 1 se exponen las principales alternativas presentes en la literatura que se han desarrollado para la clasificación de imágenes. En el capítulo 2 presenta el modelo propuesto para la implementación de un sistema de clasificación de imágenes de cáncer de piel. En los capítulos 3 y 4 se describe los algoritmos y técnicas utilizadas, se describen los experimentos realizados y se exponen los resultados obtenidos y se analiza la efectividad de estos. Finalmente, se presentan las conclusiones de la tesis y las recomendaciones para investigaciones futuras.

Capítulo 1

Estado del Arte

En las últimas décadas, los avances en potencia computacional han permitido un progreso significativo en el análisis automatizado de imágenes. Se ha pasado del análisis básico de imágenes digitales a sofisticados algoritmos capaces de identificar patrones sutiles en las imágenes de lesiones cutáneas. Los progresos en el reconocimiento de melanomas a partir de imágenes dermatológicas, han demostrado que los sistemas automatizados pueden lograr un diagnóstico comparable al de los expertos humanos [26].

1.1. Teledermatología

La teledermatología, una rama emergente de la telemedicina, ha revolucionizado el campo de la atención dermatológica. Impulsada por el auge de las tecnologías digitales a finales del siglo XX, esta modalidad se ha consolidado como una herramienta vital en el diagnóstico y manejo de afecciones cutáneas. Desde la década de 2000, la teledermatología ha permitido consultas dermatológicas remotas, ampliando significativamente el alcance de los servicios de salud [38].

La investigación de Whited et al. en 2002 fue pionera en demostrar la efectividad de esta práctica [50]. El estudio resaltó una reducción notable en los tiempos de respuesta, con una mediana de 5 días para consultas teledermatológicas, en contraste con los 28 días de los métodos tradicionales. Además, se enfatizó su utilidad en casos urgentes o semi-urgentes.

Paralelamente, la teledermatopatología ha mostrado su potencial, ofreciendo una fiabilidad comparable a la evaluación histológica tradicional [38]. Uno de los usos importantes de esta técnica fue detectar cáncer de piel. Un estudio Piccolo et al., [35] publicado en 2002, se centró en la concordancia entre los diagnósticos telepatológicos e histopatológicos convencionales, indicando contribuciones significativas en el campo de la teledermatopatología. Por término medio, el 78% de los telediagnósticos fueron

correctos (intervalo, 60%-95%), mientras que el 85% de los diagnósticos convencionales fueron correctos (intervalo, 60%-95%). Se obtuvo una concordancia diagnóstica perfecta en 7 (35%) de los 20 casos, y sólo se identificó una diferencia significativa en 1 caso.

Este avance en la teledermatología ha sentado las bases para la siguiente etapa en la medicina digital: el desarrollo de algoritmos computarizados que puedan detectar características en las imágenes, imitando el comportamiento humano, para luego a partir de estos obtener resultados.

1.2. Técnicas tempranas de clasificación de imágenes

Cuando se trata de la clasificación de imágenes, es esencial considerar que nuestro sistema visual humano (SVH) primero recibe las ondas electromagnéticas que pertenecen al espectro visible y luego las interpreta en el cerebro. Sin embargo, en el campo de la visión artificial, cuando introducimos una imagen en un ordenador, lo que se interpreta es una matriz de números generalmente en el rango de $[0,255]$ y con tres dimensiones en caso de que sea una imagen a color (RGB). Como resultado, se puede notar una gran brecha entre el significado semántico de la clase asociada a una imagen y los valores de píxeles que la componen, lo que hace que la tarea de clasificación sea compleja para sistemas artificiales. [4]

El reconocimiento de imágenes es una faceta de la inteligencia artificial que posibilita que los sistemas informáticos analicen e interpreten el contenido visual en imágenes. Esto se consigue detectando patrones y características distintivas que luego se emplean para clasificar y etiquetar objetos. Su propósito es automatizar el análisis visual, optimizando tiempo y recursos en diversas aplicaciones [36]. Esta tecnología se sustenta en algoritmos de aprendizaje automático, que entrenan a las máquinas para reconocer patrones visuales. Mediante el uso de bases de datos de imágenes etiquetadas para el entrenamiento, los algoritmos aprenden a identificar objetos y patrones. Una vez completado el entrenamiento, el modelo puede identificar automáticamente estos elementos en nuevas imágenes.

Desde finales del siglo XX, los ingenieros han dedicado esfuerzos significativos al desarrollo de técnicas y algoritmos para la categorización y reconocimiento de eventos a través de datos. Inicialmente, esto se centró en el procesamiento de texto para la clasificación automática de documentos, seguido por el tratamiento de sonidos e imágenes en diversos formatos.

Un hito importante fue la publicación de un artículo sobre reconocimiento de patrones en 1974 en *IEEE Transactions on Automatic Control* [25], evidenciando que ya en la década de 1970 se estaban implementando estas técnicas en el reconocimiento

de patrones. Los trabajos de este enfatizaron avances teóricos y experimentales significativos que impulsaron el progreso en el reconocimiento automático de patrones y el aprendizaje automático.

1.2.1. Clasificación y reconocimiento de patrones en imágenes médicas

En el ámbito de la medicina, se ha observado que la mayoría de los sistemas artificiales de diagnóstico, en un punto, toman decisiones que están cada vez menos relacionadas con la apariencia física de la imagen tal como la vería un radiólogo. En su lugar, estos sistemas se basan en los detalles del patrón matemático de las características físicas individuales de la imagen, que son extraídas por un sistema de visión artificial o un radiólogo, para tomar su decisión final. Estos patrones matemáticos han sido objeto de estudio durante décadas por científicos que han utilizado diversos métodos analíticos, incluyendo las redes neuronales [4].

Para abordar esta brecha entre la representación de la imagen y su significado, se han desarrollado varios tipos de algoritmos de clasificación. Algunos de estos algoritmos se basan en la detección de bordes, como el algoritmo de Canny [18]. Sin embargo, estos algoritmos son robustos cuando se trata de identificar una clase específica, pero si se desea clasificar una clase diferente, es necesario crear un nuevo modelo desde cero [39].

Esta serie de limitaciones restringieron su capacidad para abordar tareas complejas y desafiantes. La escasez de datos adecuados, la falta de recursos computacionales avanzados, arquitecturas simples, dificultades en el entrenamiento, generalización limitada, problemas de gradiente, falta de interpretabilidad y largos tiempos de entrenamiento fueron obstáculos clave en su desarrollo inicial.

Técnicas utilizadas en los inicios para la clasificación de patrones rompieron con algunas de las barreras de desarrollo: histograma de gradientes orientados (HOG) [14], Scale-Invariant Feature Transform (SIFT) [32], Binary Robust Independent Elementary Features (BRIEF) [10], Color Histograms [13], entre otras. Sin embargo no nos fue suficiente. Luego, con el avance de la tecnología y la llegada del machine learning el aprendizaje profundo (deep learning) también desarrolló métodos de esta índole. En la bibliografía encontramos métodos de bajo nivel como son segmentación de la imagen por niveles grises, bordes o formas, entre otras y métodos de alto nivel como clasificadores basados en redes neuronales, máquinas de soporte vectorial (SVM), árbol de decisiones, entre otros [31].

Se destaca además que las tres técnicas más usadas en la clasificación automática de imágenes son árboles de decisiones, redes neuronales y máquinas de vectores de soporte siendo las redes neuronales una de las más utilizadas en campo del aprendizaje profundo [31].

1.3. Introducción y desarrollo de machine learning en el campo médico

Las primeras aplicaciones de redes neuronales en imágenes médicas se orientaron hacia el análisis y clasificación de dichas imágenes para apoyar en diagnósticos y tratamientos.

1.3.1. Análisis de imágenes médicas

El análisis de imágenes médicas mediante redes neuronales se ha enfocado en campos de la medicina como resonancia magnética, medicina nuclear y radiología, permitiendo la identificación y clasificación de patologías o condiciones específicas. Además, estas tecnologías han encontrado aplicaciones en áreas como la oftalmología, para el diagnóstico de enfermedades oculares a partir de imágenes de retina, y en la cardiología, para la evaluación de imágenes de ecocardiogramas [4].

Resonancia Magnética: En el contexto de la esclerosis múltiple, se han explorado soluciones de segmentación basadas en redes neuronales convolucionales (CNNs) para segmentaciones rápidas y fiables de lesiones y estructuras de materia gris en imágenes de resonancia magnética multimodal [30].

Medicina Nuclear: En este campo un estudio demuestra cómo el aprendizaje profundo puede restaurar la calidad de imagen diagnóstica y mantener la precisión de la cuantificación de SUV para exploraciones PET con un conteo reducido, lo que podría aumentar la seguridad y reducir el costo de las imágenes PET [12].

Radiología: En radiología, la aplicación de la inteligencia artificial en el análisis de imágenes de cáncer, con un enfoque en radiomía y representaciones derivadas del aprendizaje profundo, y su uso para el soporte de decisiones en la gestión del cáncer [7].

Por lo que el análisis de imágenes médicas a través de redes neuronales representa un avance significativo en diversas áreas de la medicina. Las aplicaciones van desde la identificación de patologías en resonancia magnética, medicina nuclear y radiología, hasta el diagnóstico de enfermedades oculares y la evaluación cardíaca. Estas tecnologías no solo mejoran la precisión en la detección y clasificación de condiciones específicas, sino que también optimizan la eficiencia de los procesos diagnósticos y terapéuticos, destacando el papel crucial de la inteligencia artificial en el futuro de la medicina.

1.4. Datasets de cáncer de piel

En el campo de la dermatología, la generación de imágenes clínicas y dermatoscópicas es una práctica común para supervisar los cambios en las condiciones de la piel. Estas imágenes se han vuelto un recurso crucial para el avance de algoritmos de aprendizaje automático, especialmente en el desarrollo de Redes Neuronales Convolucionales (CNN). Existen varios conjuntos de datos accesibles para la investigación en este ámbito.

Das et al. [16] recoge en su estudio un conjunto de los dataset de imágenes dermatológicas más utilizados en algoritmos de clasificación que se exponen a continuación:

Entre los más destacados, se encuentra el archivo ISIC, que agrupa varios datasets de lesiones de piel clínicas y dermatoscópicas, incluidos los Desafíos ISIC, HAM10000 y BCN20000. Otros conjuntos de datos relevantes incluyen el Atlas Interactivo de Dermoscopia con 1000 ejemplos clínicos, la Biblioteca de Imágenes Dermofit con 1300 fotografías de alta resolución, el conjunto de datos PH2 con 200 imágenes dermatoscópicas, y el MED-NODE con 170 fotos clínicas.

El conjunto de datos de Asan, con 17,125 fotos clínicas, y el Hallym, con 125 fotos de casos de carcinoma basocelular, también son significativos. Además, los conjuntos de datos SD-198 y SD-260 ofrecen una amplia gama de imágenes clínicas de diversas enfermedades de la piel. Dermnet NZ y Derm7pt proporcionan colecciones extensas de fotografías clínicas, dermatoscópicas e histológicas, y The Cancer Genome Atlas presenta una de las mayores colecciones de diapositivas de lesiones cutáneas patológicas.

Entre todos estos conjuntos de datos, el HAM10000 se destaca por su amplia utilización en la investigación del cáncer de piel. Este conjunto de datos es particularmente valioso debido a su extensa colección de imágenes de lesiones de piel, que incluye una variedad de tipos de cáncer de piel. Su uso generalizado en la comunidad científica y su relevancia en estudios recientes lo convierten en el dataset ideal para el desarrollo de esta tesis. La riqueza y diversidad de las imágenes en HAM10000 proporcionan una base sólida para entrenar y evaluar algoritmos de machine learning.

1.5. Redes neuronales convolucionales en la medicina

Una red neuronal convolucional está formada por diferentes capas, entre ellas las principales son las capas convolucionales, las capas de max-pooling, y las capas completamente conectadas. La capa convolucional tiene como objetivo realizar la convolución a la imagen de entrada, para extraer sus características. Realizar una convolución a una imagen, consiste en filtrar dicha imagen utilizando una máscara o ventana. La

máscara se va desplazando por toda la imagen, multiplicándose de forma matricial [4].

En el campo de la medicina, estas facilitan la identificación de características relevantes en las imágenes médicas que pudieran ser indicativas de alguna condición médica particular. Se aprovechan de estructuras específicas de datos, como imágenes, para capturar patrones con mayor precisión, reducir la carga computacional y mejorar la generalización y la interpretabilidad.

La evolución en la clasificación de lesiones cutáneas ha estado marcada por el uso innovador de redes neuronales convolucionales (CNNs). Una revisión sistemática en 2018 por Brinker et al. [8] analizó 13 artículos que implementaban CNNs para esta tarea, destacando su alto rendimiento. Los enfoques más comunes involucraron la reutilización de CNNs ya entrenadas con grandes conjuntos de datos, optimizadas posteriormente para la clasificación específica de lesiones cutáneas. Esta metodología, aunque efectiva, enfrentó retos como la dificultad de comparar distintos métodos debido a la variabilidad en los conjuntos de datos.

Posteriormente, las técnicas de aprendizaje automático, y en particular los modelos de aprendizaje profundo, emergieron como herramientas poderosas para el análisis de imágenes médicas. Un proyecto clave fue *A Deep Learning Approach to Skin Cancer Detection in Dermoscopy Images* [5], donde se utilizó un conjunto de 3400 imágenes dermatoscópicas del HAM10000, incluyendo lesiones melanoma y no melanoma. Se implementó una red neuronal convolucional profunda que procesaba imágenes directamente, identificando características valiosas sin necesidad de segmentación previa, lo cual representó un avance significativo en la simplificación y eficacia del proceso de clasificación.

En 2022, el estudio *Skin lesion classification of dermoscopic images using machine learning and convolutional neural network* [40], publicado en Nature, utilizó un subconjunto del HAM10000 para clasificar lesiones cutáneas mediante aprendizaje automático y CNN. Este enfoque ofreció resultados prometedores en la distinción entre lesiones malignas y benignas, logrando una precisión del 95,18% con el modelo CNN. La comparación con otros algoritmos de aprendizaje automático resaltó la superioridad de los modelos CNN en términos de precisión.

Tajerian et al. [43] presentó un enfoque metodológico para mejorar el diagnóstico de la lesiones cutáneas pigmentadas utilizando imágenes dermatoscópicas del conjunto de datos HAM10000. Este conjunto de datos se utilizó para analizar lesiones cutáneas pigmentadas. El modelo obtiene los mejores resultados en la detección de lesiones de nevos melanocíticos, con una puntuación F1 de 0,93.

Adegun et al. [3] recoge un conjunto de artículos que contribuyeron al desarrollo de algoritmos además de los anteriormente mencionados:

1. Majtner et al. usaron técnicas de CNN para la extracción de características y preprocesamiento, utilizando el conjunto de datos ISIC con 900 muestras de

entrenamiento y 379 de prueba, clasificadas en benignas y malignas.

2. Vipin et al. implementaron un sistema de dos etapas, segmentación y clasificación, utilizando un conjunto de datos ISIC de 13,000 imágenes, reducido a 7,353 tras eliminar imágenes no utilizables.
3. Nasr-Esfahani et al. desarrollaron su propia CNN para preprocesar, extraer características y clasificar imágenes, con un conjunto de 170 imágenes no dermatoscópicas del UMCG, aumentado a 6,120 imágenes.
4. Attia et al. usaron una CNN completamente conectada con arquitecturas de autoencoder-decoder, alcanzando una precisión del 98% y una especificidad del 94%.
5. Mukherjee et al. desarrollaron una arquitectura CNN para la detección de lesiones malignas, logrando precisiones de 90.14% y 90.58% en los conjuntos de datos MEDNODE y Dermofit.
6. Sanketh et al. propusieron una CNN para la detección temprana de cáncer de piel, obteniendo un resultado óptimo del 98%.
7. Rahi et al. propusieron un modelo CNN con varias capas convolucionales y de agrupación máxima, logrando una precisión del 84.76% y una especificidad del 78.81%.
8. Gulati et al. emplearon redes preentrenadas como AlexNet y VGG16, obteniendo mejores resultados con VGG16 en modo de aprendizaje transferido.
9. Daghrir et al. combinaron una CNN con técnicas de aprendizaje automático clásicas, alcanzando una precisión individual del 85.5% con CNN.
10. Acosta et al. incorporaron técnicas de CNN basadas en máscaras y regiones con una estructura ResNet152 preentrenada, logrando una precisión del 90.4% y una especificidad del 92.5%.

El avance en la detección de cáncer mediante el uso de la inteligencia artificial (IA) y el aprendizaje automático (ML) ha sido significativo en las últimas décadas, el mismo, ha demostrado un rendimiento excepcional en tareas de reconocimiento de imágenes, que es fundamental en la detección del cáncer de piel. Se llevó a cabo una revisión exhaustiva para evaluar el impacto de las técnicas de aprendizaje profundo en la detección precoz, en la que se analizaron diversos resultados de investigación y se presentaron mediante herramientas, gráficos, tablas y marcos para comprender mejor las técnicas predominantes en este campo [21].

1.6. Retos y limitaciones actuales

Dificultades en la interpretación de resultados y la falta de explicabilidad de los modelos

Uno de los desafíos más significativos en el uso de algoritmos avanzados de aprendizaje automático y aprendizaje profundo es la interpretación y explicabilidad de sus resultados. Aunque estos modelos pueden lograr un alto grado de precisión, a menudo operan como "cajas negras", lo que significa que sus procesos internos y la forma en que llegan a una conclusión específica no son transparentes ni fácilmente comprensibles para los humanos. Esta falta de transparencia puede ser un obstáculo importante en el ámbito clínico, donde los profesionales de la salud necesitan comprender el razonamiento detrás de un diagnóstico para confiar y actuar según los resultados proporcionados por estos sistemas [28].

Problemas relacionados con el sobreajuste y la generalización de los modelos

Otro reto importante es el sobre-ajuste y la generalización de los modelos. El sobre-ajuste ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento, perdiendo la capacidad de generalizar a nuevos datos. Este problema es particularmente prevalente en situaciones donde los conjuntos de datos de entrenamiento son limitados o no representativos de la variabilidad real de casos clínicos. Por ejemplo, la falta de diversidad en los tipos de piel y las características de las lesiones en los conjuntos de datos puede llevar a modelos que no se desempeñan bien en poblaciones no representadas durante el entrenamiento [28].

Limitaciones de los métodos tradicionales de aprendizaje automático en diagnóstico clínico

Aunque los métodos tradicionales de aprendizaje automático, como las Máquinas de Vectores de Soporte (SVM), XGBoost y árboles de decisión, han sido útiles en la clasificación del cáncer de piel, enfrentan limitaciones significativas en el contexto clínico. Estos métodos requieren la extracción manual de características de las imágenes de enfermedades de la piel, un proceso que puede ser subjetivo y limitante. Además, la selección restringida de características puede impedir que estos algoritmos capturen la complejidad y variabilidad de las lesiones cutáneas, limitando su capacidad para generalizar a un espectro más amplio de tipos de cáncer de piel [28].

Escasez de datos y variedad limitada en categorías de cáncer de piel

La escasez de datos y la limitada variedad en las categorías de cáncer de piel presentes en los conjuntos de datos disponibles son problemas centrales en el desarrollo de algoritmos efectivos para la clasificación del melanoma. Esta limitación en los datos disponibles afecta negativamente la capacidad de los modelos para aprender y reconocer una amplia gama de manifestaciones del cáncer de piel, lo que puede resultar en un rendimiento deficiente al enfrentarse a casos menos comunes o atípicos en la práctica clínica real [28].

1.7. Conclusión del estado del arte

Los estudios revisados reflejan un claro progreso en la aplicación de CNNs y otras técnicas de aprendizaje profundo, no solo en términos de eficacia sino también en la reducción de tiempos de espera y mejora en el acceso a diagnósticos. Es notable cómo estos avances han permitido el desarrollo de sistemas de clasificación más robustos y precisos, capaces de distinguir entre lesiones cutáneas malignas y benignas con altos niveles de precisión.

En el contexto de estos avances, esta investigación aporta un valor distintivo en varios aspectos. Primero, la implementación de una arquitectura EfficientNetB5 para la clasificación de cáncer de piel, una técnica relativamente reciente y menos explorada en la literatura comparada con modelos más establecidos como AlexNet o InceptionV3. Esta elección representa un intento de equilibrar la eficiencia y precisión en un campo donde la carga computacional y la exactitud son críticas.

Además, el enfoque de esta investigación hacia el tratamiento de conjuntos de datos desequilibrados aborda una limitación significativa que ha sido un desafío persistente en estudios anteriores. Al proponer y validar métodos que manejan de manera efectiva la desproporción en las categorías de datos, se está contribuyendo a un área de necesidad crítica, mejorando potencialmente la capacidad del modelo para generalizar y funcionar eficazmente en escenarios clínicos reales.

Finalmente, esta investigación propone un análisis detallado de la interpretabilidad y la explicabilidad del modelo propuesto, un área que ha sido tradicionalmente un desafío en el campo de la inteligencia artificial. Entender cómo y por qué el modelo toma ciertas decisiones es crucial no solo para la confianza de los médicos y pacientes en estas tecnologías, sino también para identificar áreas de mejora y refinamiento en futuros trabajos.

Capítulo 2

Propuesta de solución

Este capítulo presenta una propuesta integral para abordar el desafío de la detección automática del cáncer de piel utilizando técnicas de aprendizaje profundo. La estrategia consiste en la utilización de los pesos obtenidos de una red neuronal pre-entrenada y un algoritmo de ajuste de precisión, capaz de interpretar y clasificar imágenes dermatológicas. A continuación, se esbozan los enfoques metodológicos y las herramientas tecnológicas seleccionadas para llevar a cabo este proyecto.

2.1. Desarrollo del modelo y estrategias de optimización

2.1.1. Selección del modelo y transferencia de aprendizaje

Como se explico anteriormente, el modelo consiste en utilizar los pesos obtenidos de una red neuronal convolucional pre-entrenada. Se optó por EfficientNetB1, una Red Neuronal Convolucional (CNN) pre-entrenada, como piedra angular del modelo. Esta elección se basa en la eficiencia de EfficientNet en términos de precisión y consumo de recursos computacionales [44]. Al aprovechar un modelo pre-entrenado, se utilizan pesos obtenidos de extensos conjuntos de datos de imágenes, lo que facilita la adaptación a nuestro conjunto de datos específico. La omisión de la última capa de clasificación permite una personalización más profunda, adaptando el modelo para identificar con precisión las variadas presentaciones de tumores cutáneos presentes en el HAM1000.

Arquitectura del modelo EfficientNet

La familia de arquitecturas EfficientNet, desarrollada por los autores en [44], surgió con el objetivo de hallar un método adecuado para escalar las CNNs de manera que

mejoraran tanto en precisión (i.e., rendimiento del modelo) como en eficiencia (es decir, en términos de parámetros del modelo y FLOPS). Estos autores propusieron un método de escalado compuesto que utiliza un conjunto fijo de coeficientes para escalar de manera uniforme el ancho, la profundidad y la resolución de la red. El método les permitió desarrollar una arquitectura de CNN eficiente, a la que denominaron EfficientNet B0. Posteriormente, crearon las variantes EfficientNets B1-B7 escalando la red base (EfficientNet B0).

Mientras que la arquitectura EfficientNet B0 tiene 5,3 millones de parámetros y acepta imágenes de entrada de 224×224 , EfficientNet B7 cuenta con 66 millones de parámetros y acepta imágenes de 600×600 [45].

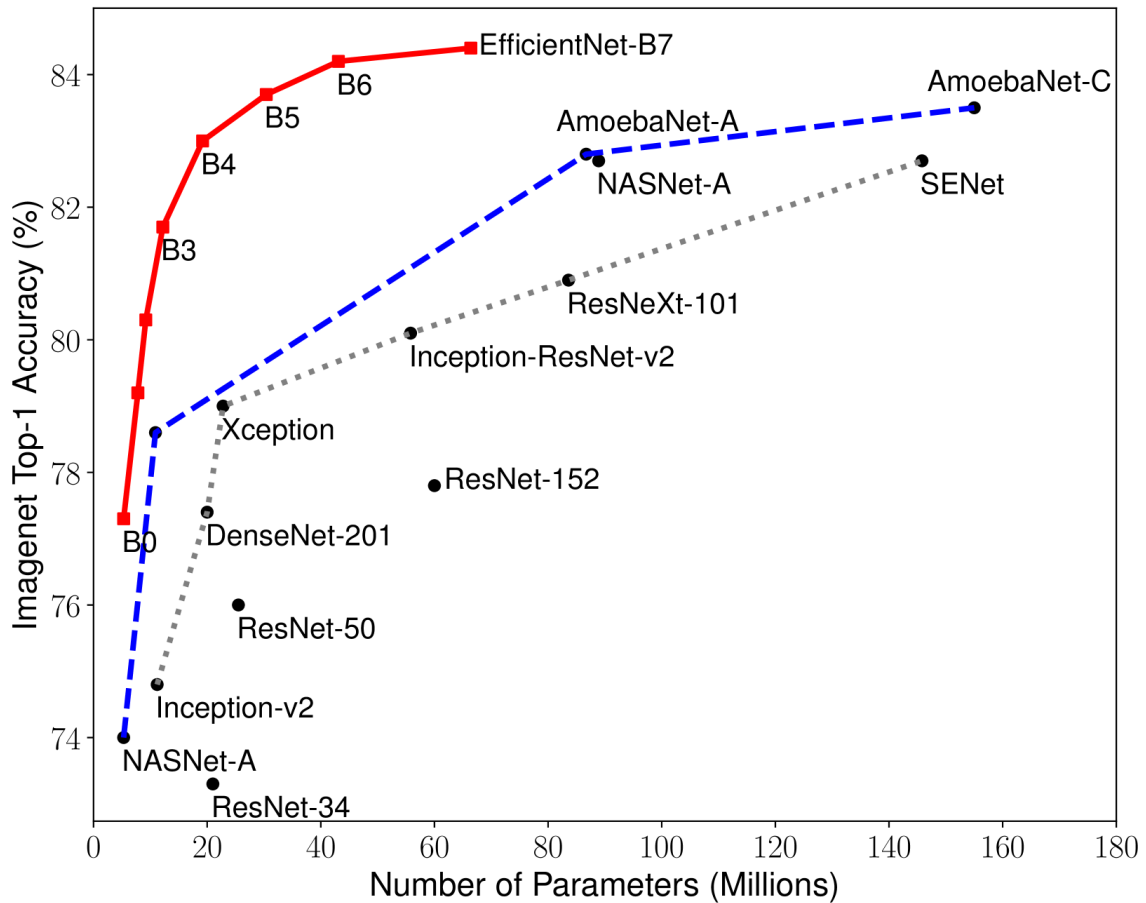


Figura 2.1: Estadísticas del rendimiento de los modelos de EfficientNet

EfficientNetB1

En relación con el dataset HAM1000 [15], la implementación del EfficientNetB1 puede ser particularmente beneficiosa para el análisis de datos. EfficientNetB1, entre las variantes de la serie EfficientNet, se encuentra en un punto medio en términos de complejidad y tamaño, ofreciendo un equilibrio entre precisión y eficiencia computacional. Dado que el HAM1000 es un conjunto de datos de imágenes dermatoscópicas que requiere una alta precisión en la identificación y clasificación de lesiones cutáneas, la utilización de EfficientNetB5 podría proporcionar una precisión y eficiencia aceptable en términos de recursos computacionales.

2.1.2. Preprocesamiento de datos y balanceo de clases

El proceso de preprocesamiento incluye el re-dimensionamiento y la normalización de las imágenes, ajustándolas a los requisitos de entrada de EfficientNet. Dada la naturaleza del dataset HAM1000, que a menudo muestra un desequilibrio en la representación de clases, se puso especial énfasis en el balanceo de clases. Este paso es crucial para evitar sesgos en el modelo y asegurar que todas las categorías de tumores sean tratadas con igual importancia. Además, se empleó la técnica de aumento de datos para enriquecer el conjunto de entrenamiento, mejorando la capacidad del modelo para generalizar a partir de datos variados y no vistos anteriormente.

2.1.3. Arquitectura del Modelo y Regularización

Para fortalecer la arquitectura del modelo, se añadieron capas adicionales, incluyendo Dropout y regularizadores L1 y L2, esenciales para combatir el sobre-ajuste. Una capa densa personalizada fue incorporada para facilitar la clasificación precisa de múltiples tipos de tumores. La compilación del modelo se realizó con un enfoque en la clasificación multi-clase, utilizando la pérdida de entropía cruzada categórica y un optimizador adecuado, seleccionados por su efectividad en tareas similares.

2.1.4. Ajuste dinámico del learning rate

Un elemento innovador del entrenamiento fue la implementación de un callback personalizado para el ajuste dinámico del learning rate. Esta estrategia permite ajustes adaptativos del learning rate basados en el rendimiento del modelo, optimizando la eficiencia del entrenamiento y evitando el estancamiento en mínimos locales del espacio de búsqueda.

2.1.5. Diversidad en propuestas de soluciones

El proyecto explora diferentes enfoques tanto en el preprocesamiento de datos como en la elección de optimizadores. Esta variedad en las técnicas utilizadas refleja un esfuerzo por abordar el problema desde múltiples perspectivas, aumentando así las probabilidades de encontrar la solución más eficaz para el desafío específico del dataset HAM1000.

Capítulo 3

Detalles de implementación y experimentos

Para la implementación del modelo y con el objetivo de lograr mejor precisión y generalización, se llevaron a cabo varios experimentos. La metodología de los experimentos fue la misma. Dado que el conjunto de datos utilizado para la validación del modelo está desbalanceado la cantidad de imágenes por clase, en los distintos experimentos se modificó la distribución de datos para aprendizaje, test y validación. Se exponen durante el capítulo 2 de los experimentos que mostraron tener mejores resultados en cuanto a precisión y recobrado.

3.1. Enfoque de Implementación

La implementación seguirá un enfoque modular y escalable. Comenzará con la selección del modelo y su personalización para el contexto específico del cáncer de piel. El preprocesamiento de datos y el balanceo de clases asegurarán la calidad del entrenamiento y la generalización del modelo. La arquitectura del modelo será iterativamente refinada y evaluada, utilizando un algoritmo propio de ajuste de precisión. Además se utilizan capas adicionales de ajuste y optimización para lograr mejores resultados.

3.1.1. Herramientas y Tecnologías

La implementación del modelo se llevará a cabo utilizando las siguientes herramientas:

- **Lenguajes de Programación:** Python será utilizado por su rica biblioteca de paquetes de aprendizaje automático, incluyendo TensorFlow y Keras para la construcción y entrenamiento del modelo.

- **Frameworks de Aprendizaje Profundo:** TensorFlow y Keras proporcionan las funcionalidades necesarias para diseñar, entrenar y validar modelos de aprendizaje profundo con alta eficiencia y flexibilidad.
- **Técnicas de Preprocesamiento:** Herramientas para la normalización de imágenes, el aumento de datos, y el balanceo de clases serán utilizadas para preparar el dataset para el entrenamiento del modelo.
- **Optimización y Regularización:** Se integrarán técnicas como el ajuste dinámico del learning rate y la regularización L1 y L2 para optimizar el rendimiento del modelo y prevenir el sobre-ajuste.
- **Hardware y Recursos Computacionales:** Se utilizó Google Collab para acelerar el proceso de entrenamiento del modelo, permitiendo la experimentación con diferentes hiper-parámetros y arquitecturas de forma eficiente.

3.2. Preparación y carga de datos

3.2.1. Dataset

El dataset utilizado como medio de aprendizaje para este proyecto es HAM1000-segmentation-and-classification [15]. Este conjunto de datos, acrónimo de Human Against Machine (Humano contra máquina) con 10000 imágenes de entrenamiento, es una amplia colección de imágenes dermatoscópicas. En concreto, contiene 10015 imágenes del archivo ISIC, que inicialmente formaban parte de un conjunto de entrenamiento creado por ISIC [47] con la siguiente distribución.

| Categoría Diagnóstica | Número de Imágenes | Porcentaje |
|-------------------------------|--------------------|------------|
| Melanocytic nevi | 6705 | 66.95 % |
| Melanoma | 1113 | 11.11 % |
| Benign keratosis-like lesions | 1099 | 10.97 % |
| Basal cell carcinoma | 514 | 5.13 % |
| Actinic keratoses | 327 | 3.27 % |
| Vascular lesions | 142 | 1.42 % |
| Dermatofibroma | 115 | 1.15 % |

Los datos utilizados como medio de aprendizaje para este proyecto son imágenes y metadatos. El dataset HAM10000 contiene imágenes tomadas de varios tipos de cáncer de piel y un archivo de metadatos que contienen información relacionada con cada imagen en formato *one hot encoding*. Esta es una técnica de procesamiento de datos en la cual cada valor categórico se representa mediante un vector binario cuyo tamaño corresponde al número de categorías posibles. En dicho vector, todos los elementos son cero, salvo el correspondiente a la categoría del valor, que es uno [34].

3.2.2. Transformación de datos

Para optimizar la eficiencia del algoritmo, se procesan las imágenes realizando diversas modificaciones, que incluyen: eliminación de cabello, luces y sombras, división de canales y aplicación de un leve desenfoque.

Los metadatos asociados a la clasificación, etiquetados con método mencionado, fueron convertidos a un formato *categorico* [41] para su procesamiento. Se define *categorico* como el formato en el que a cada elemento se le asigna el nombre de una categoría como propiedad.

3.2.3. Modelación y división del conjunto de datos

Los datos se modelan a partir de un dataframe de Pandas [17]. Estos son divididos en 3 conjuntos: Entrenamiento, Validación y Prueba, utilizando un enfoque simple de división de datos en porcentaje. Estas divisiones son necesarias para que el modelo pueda aprender y ser evaluado correctamente.

Inicialmente, el conjunto de datos, fue dividido en dos subconjuntos: *train*, que se destinó para el entrenamiento, y *dummy*, que fue utilizado como una combinación temporal para los conjuntos de validación y prueba. Luego el segundo conjunto fue separado en *valid*, destinado a la validación y *test*, utilizado para las pruebas.

3.2.4. Generadores de datos y preprocesamiento

Uno de los principales problemas del dataset, es el desequilibrio en la representación de las clases, un problema habitual en los conjuntos de datos médicos en los que algunas enfermedades son más raras que otras. Para solucionar este problema, el conjunto de datos se equilibra meticulosamente limitando el número máximo de muestras por clase (300). Esto garantiza que el modelo no esté sesgado hacia las clases más comunes y pueda generalizar mejor entre varios tipos de lesiones cutáneas.

3.2.5. Aumento y división de datos

Para la carga y procesamiento de imágenes se utilizó la clase `ImageDataGenerator` de Keras [27]. Esta clase es una parte integral de la biblioteca Keras y proporciona una forma eficiente de manipular imágenes para tareas de aprendizaje automático. Su principal función es facilitar la creación de lotes de imágenes que se utilizan durante el entrenamiento y la evaluación de modelos de Machine Learning, especialmente en el contexto de redes neuronales.

Se inicializa con varias transformaciones (como rotación, desplazamiento y zoom) aplicadas automáticamente a las imágenes a medida que se cargan, para devolver lotes de imágenes listos para el entrenamiento o la evaluación del modelo [6].

Tabla 3.1: Parámetros de aumento de datos

| Parámetro | Descripción | Valor |
|--------------------|--|-----------|
| rotation range | Rango de rotación | 20 grados |
| width shift range | Rango de desplazamiento horizontal | 20 % |
| height shift range | Rango de desplazamiento vertical | 20 % |
| shear range | Rango de corte | 20 % |
| zoom range | Rango de zoom | 20 % |
| horizontal flip | Activación de volteo horizontal | verdadero |
| fill mode | Modo de relleno para manejar los píxeles faltantes | nearest |

3.3. Diseño y entrenamiento del modelo

Se utiliza la arquitectura de red neural denominada *EfficientNetB1* [23]. Esta arquitectura forma parte de la familia EfficientNet, que está diseñada para proporcionar alta precisión mientras se mantiene un tamaño de modelo y una complejidad computacional eficientes.

Para aprovechar los conocimientos previos y acelerar el entrenamiento, se carga el modelo *EfficientNetB1* pre-entrenado con los pesos obtenidos al procesar el dataset *ImageNet*. El mismo consiste en una amplia base de datos de imágenes utilizada comúnmente para entrenamiento y *benchmarking* en tareas de visión por computadora. Utilizar este modelo pre-entrenado permite aprovechar las características que ya ha aprendido de este amplio conjunto de datos, facilitando su adaptación al dataset HAM10000.

Se utiliza el *EfficientNetB1* sin incluir su capa superior, con el objetivo de añadir y personalizar capas adicionales. Tras obtener la salida del modelo base, se aplica una serie de transformaciones, incluyendo normalización por lotes, una capa densa con regularizaciones, una técnica de *Dropout* para prevenir el sobre-ajuste y una capa de optimización.

3.3.1. Capas adicionales y regularización

Para ajustar el modelo de EfficientNetB1 a nuestras necesidades, se añaden capas adicionales:

Normalización por lotes

Para mejorar la estabilidad y eficiencia del modelo EfficientNetB1, se incorpora una capa de Normalización por lotes. La normalización por lotes es crucial para estabilizar las *activaciones* a lo largo de las capas del modelo. Al estandarizar las activaciones por lote, mejora la eficiencia del entrenamiento y permite el uso de tasas

de aprendizaje más altas. Esta técnica contribuye significativamente a la regularización del modelo, reduciendo el riesgo de sobre-ajuste y facilitando un aprendizaje más rápido y estable.

Esta capa se define con un *eje de normalización* establecido en -1 , lo que indica que la normalización se aplica a lo largo del último eje en el tensor de entrada. Además, se configura un *momentum* de $0,99$ y un valor de *epsilon* de $0,001$. El alto valor de momentum ayuda a mantener la estabilidad de las medias y varianzas móviles a lo largo del entrenamiento, mientras que el pequeño valor de epsilon evita divisiones por cero, asegurando así cálculos numéricos estables [37].

Capa densa

Se integra una capa densa con 256 neuronas, que juega un papel clave en la síntesis de las características aprendidas por el modelo. Se utiliza un regularizador *L2* con un *lambda* de $0,016$ para los pesos de la capa, lo que ayuda a penalizar y controlar el tamaño de los pesos, reduciendo así el riesgo de sobre-ajuste. Además, tanto el regularizador de actividad como el regularizador de bias se configuran con un regularizador *L1* con un *lambda* de $0,006$. Este enfoque impone una penalización en los pesos y los sesgos, promoviendo un modelo más simple y disperso. La función de activación utilizada es *ReLU*, conocida por su eficacia en la introducción de no linealidad en el modelo, lo que permite aprender relaciones complejas entre las características [11].

Dropout

En la arquitectura del modelo, se integra una capa de dropout para aumentar la robustez y prevenir el sobre-ajuste. Esta capa se configura con una tasa de desactivación del 45% , lo que significa que, durante el entrenamiento, el 45% de las neuronas se desactivarán aleatoriamente en cada paso.

En una primera iteración del algoritmo tuvo una tasa de desactivación más baja. Luego de varias iteraciones se concluyó que se necesitaba un algoritmo de clasificación que estudiara más a detalle la data fomentando así una mejor generalización y reduciendo el riesgo de sobre-ajuste en el proceso de aprendizaje. Basándonos en esto aumentamos la tasa y obtuvimos mejores resultados.

Para asegurar la reproducibilidad, se establece una semilla(seed) 123. Esta introducción de aleatoriedad ayuda a que el modelo no dependa excesivamente de ninguna característica o neurona específica [22].

Capa de salida

La capa de salida utiliza una activación *softmax* para transformar las salidas del modelo en probabilidades de pertenencia a cada clase. Esto, clasifica las entradas en categorías distintas, proporcionando probabilidades para cada clase, lo cual es esencial en la clasificación multi-clase.

3.3.2. Optimización

Para el proceso de entrenamiento, se utiliza el optimizador *Adamax* [2]. Este es una variante del conocido optimizador Adam, que combina las ventajas de los métodos adaptativos de tasa de aprendizaje con una implementación más robusta en entornos con gradientes dispersos, lo cual es común en imágenes médicas.

La función de pérdida elegida es la *categorical crossentropy* [41], idónea para problemas de clasificación multi-clase. Se configura el modelo para minimizarla y se rastrea la precisión como métrica principal.

3.3.3. Ajuste dinámico de la tasa de aprendizaje

Un componente innovador de nuestro enfoque es el uso de un Callback personalizado para el ajuste dinámico del learning rate [46], basado en la precisión del entrenamiento y la pérdida de validación. Este mecanismo adapta el learning rate durante el entrenamiento, reduciéndolo si el modelo no mejora a un ritmo esperado. Este ajuste es vital en la navegación de superficies de pérdida complejas, aumentando la probabilidad de que el modelo evite mínimos locales subóptimos y converja hacia soluciones más efectivas.

3.4. Experimentos

3.4.1. Experimento 1: Evaluación de la eficiencia de la división asimétrica de datos en la clasificación de imágenes de cáncer de piel

Este experimento se centra en una división de datos altamente asimétrica, con un enfoque predominante en el conjunto de entrenamiento. La técnica de *dummy split* se emplea para mantener proporciones consistentes entre los conjuntos de validación y prueba. Esta metodología es relevante para evaluar el impacto de un extenso conjunto de entrenamiento en la precisión y el rendimiento del modelo.

La distribución de datos fue la siguiente:

1. El 95% de los datos se destinan al conjunto de entrenamiento.
2. El 2.5% de los datos restantes se destinan al conjunto de validación.
3. El 2.5% restante se destina al conjunto de pruebas.

Además se mezclan aleatoriamente los datos y se utiliza una variable fija para garantizar que la división sea reproducible. Aquí se utiliza *dummy split* para mantener la proporción deseada entre validación y prueba. Por lo que este experimento tiene 9514 datos de entrenamiento, 251 de test y 250 de validación.

Tabla 3.2: Experimento 1: Distribución de imágenes de cáncer de piel en los conjuntos de entrenamiento, test y validación

| Diagnostic category | Training | Validation | Testing |
|---------------------|----------|------------|---------|
| NV | 300 | 158 | 163 |
| MEL | 300 | 25 | 35 |
| BKL | 300 | 34 | 30 |
| DF | 115 | 5 | 3 |
| AKIEC | 300 | 11 | 7 |
| BCC | 300 | 16 | 10 |
| VASC | 142 | 1 | 3 |

Generadores de datos y preprocesamiento

Uno de los principales problemas del dataset, es el desequilibrio en la representación de las clases, un problema habitual en los conjuntos de datos médicos en los que algunas enfermedades son más raras que otras. Para solucionar este problema, el conjunto de datos se equilibra meticulosamente limitando el número máximo de muestras por clase (300). Esto garantiza que el modelo no esté sesgado hacia las clases más comunes y pueda generalizar mejor entre varios tipos de lesiones cutáneas.

Se establece para este un tamaño objetivo de muestras por clase (300 en este caso), y se utiliza un bucle para iterar a través de cada clase única. Se realiza un re-muestreo con reemplazo para clases con un número de muestras menor al objetivo (300), y sin reemplazo para clases con un número igual o mayor al tamaño objetivo.

Como se evidencia anteriormente, al separar la data en clases el conjunto se mantiene desbalanceado. Se hace necesario aplicar un método llamado *Class Weighting* [48]. Este método se refiere a la asignación de pesos diferenciados a cada clase durante el proceso de entrenamiento del modelo, con el objetivo de reforzar la señal de entrenamiento de las clases menos representadas, o sea, las clases/categorías con menos cantidad de datos:

Tabla 3.3: Distribución de muestras por categoría después del sobre-muestreo

| Diagnostic category | Sampling |
|---------------------|----------|
| AKIEC | 300 |
| BCC | 300 |
| BKL | 300 |
| DF | 115 |
| MEL | 300 |
| NV | 300 |
| VASC | 142 |

Tabla 3.4: Distribución de muestras con peso asignado

| Diagnostic category | Sampling | Weighting |
|---------------------|----------|-----------|
| AKIEC | 300 | 1.00 |
| BCC | 300 | 1.00 |
| BKL | 300 | 1.00 |
| DF | 115 | 2.60 |
| MEL | 300 | 1.00 |
| NV | 300 | 1.00 |
| VASC | 142 | 2.11 |

3.4.2. Experimento 2: Análisis de la estratificación de datos en la clasificación de imágenes de cáncer de piel

Descripción: Este experimento explora la estratificación de datos para mantener una distribución uniforme de etiquetas en cada conjunto de datos. La proporción de los conjuntos de datos es más equilibrada en comparación con el Experimento 1, lo que ofrece insights sobre la importancia de la distribución equitativa de datos en el entrenamiento y evaluación de modelos.

La distribución de datos fue la siguiente:

1. El 70% de los datos se destinan al conjunto de entrenamiento.
2. El 15% de los datos restantes se destinan al conjunto de validación.
3. El 15% restante se destina al conjunto de pruebas.

Se utiliza *stratify* en ambas divisiones para mantener la distribución de etiquetas *label* en cada conjunto. Se calcula la proporción del conjunto de prueba sobre la suma del conjunto de test y el de validación.

Luego este quedaría distribuido en 7010 datos de entrenamiento, 1503 de test y 1502 de validación.

Tabla 3.5: Experimento 2: Distribución de imágenes de cáncer de piel en los conjuntos de entrenamiento, test y validación

| Diagnostic category | Training | Validation | Testing |
|---------------------|----------|------------|---------|
| NV | 300 | 1006 | 1006 |
| MEL | 300 | 167 | 167 |
| BKL | 300 | 165 | 165 |
| DF | 300 | 17 | 17 |
| AKIEC | 300 | 49 | 49 |
| BCC | 300 | 77 | 77 |
| VASC | 300 | 21 | 22 |

Generadores de datos y preprocesamiento

Se establece también un tamaño objetivo de muestras por clase (300). Se utiliza la función *groupby* para agrupar el *dataFrame* por la etiqueta de clase. Se itera sobre cada grupo, y se realiza un re-muestreo con reemplazo para grupos menores al tamaño deseado y sin reemplazo para los grupos que ya alcanzan o superan el tamaño deseado.

En este, a diferencia del primero, en cada clase se alcanza la misma cantidad de muestras, por lo que no es necesario aplicar *Class Weighting*.

Tabla 3.6: Distribución de muestras por categoría después del sobre-muestreo

| Diagnostic category | Sampling |
|---------------------|----------|
| AKIEC | 300 |
| BCC | 300 |
| BKL | 300 |
| DF | 300 |
| MEL | 300 |
| NV | 300 |
| VASC | 300 |

Capítulo 4

Resultados

En esta sección presentamos los resultados obtenidos. Nuestro enfoque se centró en la exploración y optimización de dos pipelines de procesamiento y clasificación de imágenes. Los experimentos se estructuraron con el objetivo de evaluar mejoras de precisión en el modelo utilizando distintas aproximaciones de división y normalización de datos en el contexto del procesamiento de imágenes.

El modelo con mayor eficiencia luego de varios ajustes tuvo una eficacia de validación cercana al 84 %. Los resultados son un paso significativo hacia la automatización del diagnóstico del cáncer de piel, que tradicionalmente ha dependido de la inspección visual por parte de expertos humanos.

En cada sección se muestran algunas métricas que muestran el desarrollo del modelo, con los siguientes parámetros:

4.1. Experimento 1: Evaluación de la eficiencia de la división asimétrica de datos en la clasificación de imágenes de cáncer de piel

Estadísticas básicas

Estos resultados de la tabla siguiente corresponden a la evaluación del modelo a lo largo de 40 epochs (iteraciones) de entrenamiento.

Aquí vemos que el entrenamiento se planeó para 40 epochs, pero se detuvo en el epoch 34, después de tres ajustes de la tasa de aprendizaje sin mejoras en la pérdida de validación. Esto ocurre dado que el modelo no estaba mejorando su capacidad de generalizar a datos no vistos y que continuar el entrenamiento probablemente habría resultado en un gasto innecesario de recursos y posiblemente en un mayor sobre-ajuste.

Tabla 4.1: Descripción de términos clave en el entrenamiento de modelos de aprendizaje automático.

| Término | Descripción |
|---------------------------------|---|
| Epoch | Es una iteración completa sobre todo el conjunto de datos de entrenamiento. |
| Loss (Pérdida) | Es una medida de cuán bien el modelo está realizando sus predicciones. Los valores decrecientes indican una mejora en el aprendizaje. |
| Accuracy (Precisión) | Muestra el porcentaje de etiquetas que el modelo predice correctamente para el conjunto de entrenamiento. |
| V loss (Pérdida de Validación) | Es similar a la pérdida, pero se calcula sobre un conjunto de datos que no se utiliza para el entrenamiento. |
| V acc (Precisión de Validación) | Muestra el porcentaje de etiquetas que el modelo predice correctamente para el conjunto de datos de validación. |
| LR (Learning Rate) | La tasa de aprendizaje dicta cuánto se ajustan los pesos del modelo en cada actualización. |
| Next LR (Próxima Learning Rate) | Indica la próxima tasa de aprendizaje planificada. La adaptación de la tasa de aprendizaje puede ayudar a evitar el estancamiento y mejorar la convergencia. |
| Monitor | Muestra la métrica que se está utilizando para monitorizar el rendimiento del modelo. Cambia de <i>accuracy</i> a <i>val loss</i> , lo que probablemente indica que el cambio se hizo para evitar el sobreajuste. |
| Duration (Duración) | Tiempo que tardó cada epoch en completarse. Importante para evaluar la eficiencia del entrenamiento. |

En la tabla se observa que la precisión de entrenamiento aumenta con cada epoch, la pérdida de entrenamiento disminuye consistentemente, la tasa de aprendizaje permanece constante al principio y luego disminuye para afinar el entrenamiento a medida que el modelo comienza a converger, todo lo anterior es indicador que el modelo está entrenando de forma correcta.

En el monitor se muestra que cambia de *accuracy* a *val loss*, este cambio se hizo para evitar el sobreajuste.

Estadísticas de aprendizaje

La imagen anterior muestra dos gráficos, uno de pérdida y otro de precisión respectivamente, a lo largo de los epochs de entrenamiento y validación. El primero muestra una disminución constante de la pérdida tanto en el entrenamiento como en la validación, lo cual es indicativo de un buen aprendizaje. La pérdida de validación

| E | Loss | Acc | V loss | V acc | LR | M | Batch |
|-----|-------|--------|---------|--------|-----------|----------|-------|
| 1 | 9.587 | 40.581 | 8.95658 | 56.800 | 10^{-2} | accuracy | 85.25 |
| 2 | 7.798 | 67.615 | 7.67235 | 66.800 | 10^{-2} | accuracy | 21.72 |
| 3 | 6.884 | 79.340 | 6.96014 | 69.600 | 10^{-2} | accuracy | 22.56 |
| 4 | 6.214 | 87.365 | 6.35865 | 71.200 | 10^{-2} | accuracy | 25.81 |
| 5 | 5.646 | 91.690 | 5.94812 | 75.200 | 10^{-2} | val_loss | 23.08 |
| 6 | 5.172 | 92.999 | 5.44954 | 76.800 | 10^{-2} | val_loss | 23.23 |
| 7 | 4.735 | 94.479 | 5.06016 | 76.400 | 10^{-2} | val_loss | 23.19 |
| 8 | 4.334 | 96.528 | 4.73837 | 76.800 | 10^{-2} | val_loss | 22.90 |
| 9 | 3.969 | 97.211 | 4.33689 | 77.200 | 10^{-2} | val_loss | 22.74 |
| 10 | 3.631 | 98.008 | 4.15826 | 74.000 | 10^{-2} | val_loss | 22.56 |
| 11 | 3.315 | 98.406 | 3.89153 | 73.600 | 10^{-2} | val_loss | 23.11 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 34 | 0.627 | 99.886 | 1.24470 | 76.800 | 0.00013 | val_loss | 23.30 |

Figura 4.1: Estadísticas básicas del modelo EfficientNetB1.

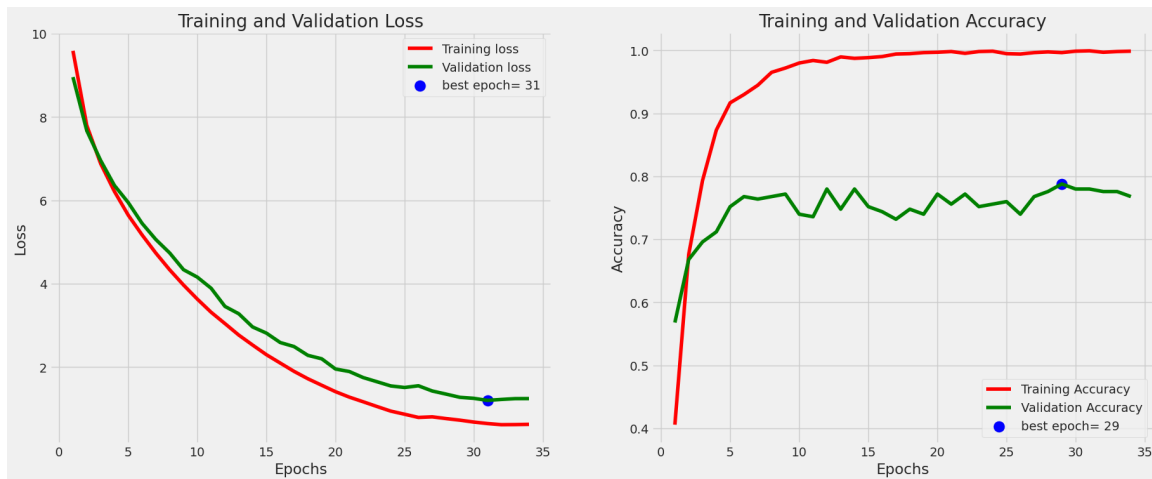


Figura 4.2: Estadísticas de aprendizaje a lo largo del proceso de entrenamiento.

tiene un mínimo en el epoch 31, lo que sugiere que este podría ser el mejor modelo para evitar el sobre-ajuste. A partir de este punto, si la pérdida de validación comenzase a aumentar mientras la pérdida de entrenamiento sigue disminuyendo, indicaría sobre-ajuste.

En el segundo se aprecia como la precisión de entrenamiento aumenta rápidamente y luego se estabiliza cerca del 100%, mientras que la precisión de validación también aumenta pero con cierta variabilidad entre los epochs. El mejor epoch según

la precisión de validación es el 29. Esta diferencia entre la precisión de entrenamiento y de validación sugiere que puede haber variabilidad en los datos de validación.

Estadísticas de eficacia

| | | Confusion Matrix | | | | | | |
|--------|-------|------------------|-----|-----|----|-----|-----|------|
| Actual | AKIEC | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| | BCC | 0 | 10 | 0 | 0 | 0 | 0 | 0 |
| | BKL | 2 | 2 | 19 | 0 | 5 | 1 | 1 |
| | DF | 0 | 0 | 0 | 3 | 0 | 0 | 0 |
| | MEL | 2 | 3 | 1 | 0 | 28 | 1 | 0 |
| | NV | 4 | 2 | 9 | 6 | 7 | 134 | 1 |
| | VASC | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| | | AKIEC | BCC | BKL | DF | MEL | NV | VASC |
| | | Predicted | | | | | | |

Figura 4.3: Estadísticas de eficacia del modelo al estimar los resultados en el conjunto de pruebas

La matriz de confusión proporciona información valiosa sobre el rendimiento del modelo en relación de Actual/Predicho, en términos de su capacidad para clasificar correctamente cada una de las siete clases de cáncer de piel. La diagonal principal de la matriz representa los verdaderos positivos (TP), el número de casos en los que el modelo ha predicho correctamente la clase correspondiente. Los valores fuera de la diagonal principal indican errores de clasificación.

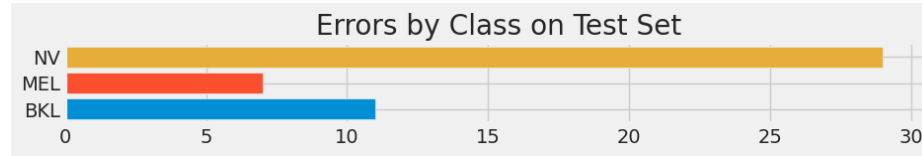


Figura 4.4: Gráfico de errores por clase en el conjunto de pruebas

La clase 'NV' muestra una alta cantidad de clasificaciones correctas (134), pero también tiene errores significativos al ser confundida con otras clases (como 'BKL' y 'MEL'). La clase 'AKIEC' resultó la más difícil de predecir, con la mayoría de las muestras mal clasificadas. Esto puede deberse a la similitud entre las características de estas clases, pero sobre todo al desbalance en el conjunto de datos.

Cada barra representa el número de muestras de una clase particular que fueron mal clasificadas.

Informe de clasificación

Tabla 4.2: Informe de clasificación para el Experimento 1

| Categoría Diagnóstica | Precisión | Recall | F1-Score | Soporte |
|-----------------------|-----------|--------|----------|---------|
| AKIEC | 0.47 | 1.00 | 0.64 | 7 |
| BCC | 0.59 | 1.00 | 0.74 | 10 |
| BKL | 0.66 | 0.63 | 0.64 | 30 |
| DF | 0.33 | 1.00 | 0.50 | 3 |
| MEL | 0.70 | 0.80 | 0.75 | 35 |
| NV | 0.99 | 0.82 | 0.90 | 163 |
| VASC | 0.60 | 1.00 | 0.75 | 3 |
| Accuracy | | | 0.81 | 251 |
| Macro Avg | 0.62 | 0.89 | 0.70 | 251 |
| Weighted Avg | 0.86 | 0.81 | 0.83 | 251 |

Resultados Generales: La precisión global del modelo fue del 81%, con una media ponderada de precisión del 86%, recall del 81%, y una puntuación F1 del 83%. Resultados por Categoría Diagnóstica: Describir la precisión, el recall, y la puntuación F1 para cada categoría diagnóstica, destacando que 'NV' tuvo la precisión más alta (0.99) y 'AKIEC' la más baja (0.47). El recall fue perfecto (1.00) para 'AKIEC', 'BCC', 'DF', y 'VASC', lo que indica que todas las muestras verdaderas fueron identificadas correctamente, pero la precisión varió, sugiriendo un posible desequilibrio en la identificación de falsos positivos.

4.2. Experimento 2: Análisis de la estratificación de datos en la clasificación de imágenes de cáncer de piel

Estadísticas básicas

Estos resultados de la tabla siguiente corresponden a la evaluación del modelo a lo largo de 40 epochs (o iteraciones) de entrenamiento. Cada fila representa una época y se presentan las siguientes métricas:

| E | Loss | Acc | V loss | V acc | LR | M | Batch |
|-----|-------|--------|---------|--------|-----------|----------|-------|
| 1 | 9.587 | 40.581 | 8.95658 | 56.800 | 10^{-2} | accuracy | 85.25 |
| 2 | 7.798 | 67.615 | 7.67235 | 66.800 | 10^{-2} | accuracy | 21.72 |
| 3 | 6.884 | 79.340 | 6.96014 | 69.600 | 10^{-2} | accuracy | 22.56 |
| 4 | 6.214 | 87.365 | 6.35865 | 71.200 | 10^{-2} | accuracy | 25.81 |
| 5 | 5.646 | 91.690 | 5.94812 | 75.200 | 10^{-2} | val_loss | 23.08 |
| 6 | 5.172 | 92.999 | 5.44954 | 76.800 | 10^{-2} | val_loss | 23.23 |
| 7 | 4.735 | 94.479 | 5.06016 | 76.400 | 10^{-2} | val_loss | 23.19 |
| 8 | 4.334 | 96.528 | 4.73837 | 76.800 | 10^{-2} | val_loss | 22.90 |
| 9 | 3.969 | 97.211 | 4.33689 | 77.200 | 10^{-2} | val_loss | 22.74 |
| 10 | 3.631 | 98.008 | 4.15826 | 74.000 | 10^{-2} | val_loss | 22.56 |
| 11 | 3.315 | 98.406 | 3.89153 | 73.600 | 10^{-2} | val_loss | 23.11 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 34 | 0.627 | 99.886 | 1.24470 | 76.800 | 0.00013 | val_loss | 23.30 |

Figura 4.5: Estadísticas básicas de algunas iteraciones del modelo EfficientNetB1.

En general, los resultados muestran que el modelo mejora con el tiempo, ya que la pérdida disminuye y la precisión aumenta tanto en el conjunto de datos de entrenamiento como en el de validación.

Estadísticas de aprendizaje

La época con la mejor pérdida de validación y la mejor precisión de validación no coinciden, sugiere un trade-off entre la optimización de la pérdida y la maximización de la precisión. La volatilidad en la precisión de validación sugiere que el modelo puede beneficiarse de un ajuste en el tamaño del lote para suavizar las actualizaciones de los pesos y mejorar la estabilidad del modelo.

Dado lo anterior, se hizo una última iteración del modelo aumentando el volumen de muestras por clase a 500. Estos fueron los resultados. Se obtuvo una efectividad

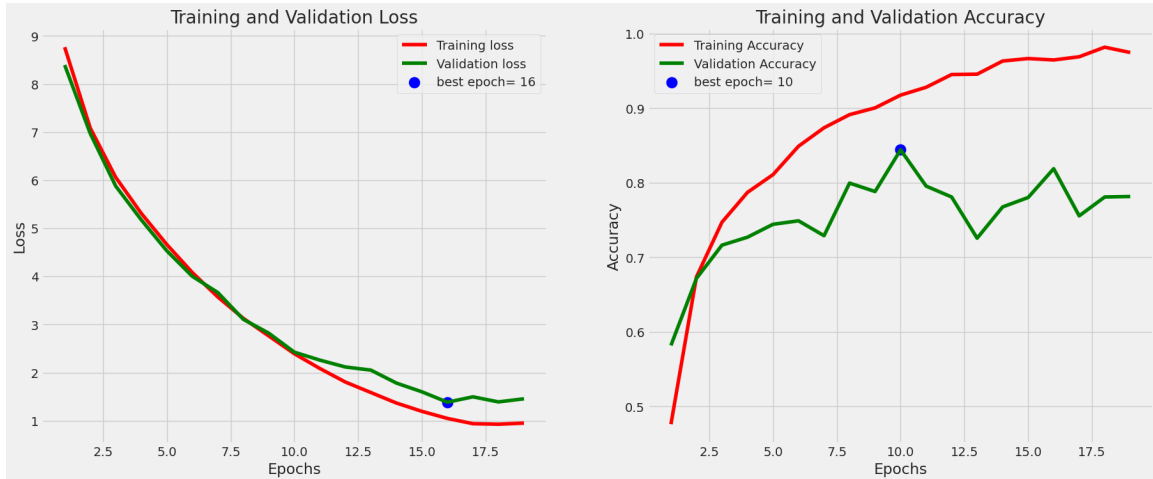


Figura 4.6: Estadísticas de aprendizaje a lo largo del proceso de entrenamiento.

de aproximadamente 87% (86.83%).



Figura 4.7: Estadísticas de aprendizaje a lo largo del proceso de entrenamiento.

La disminución continua en la pérdida de entrenamiento y la tendencia ascendente en la precisión de entrenamiento sugieren que el modelo está aprendiendo de manera efectiva y consistente. La pérdida de validación que converge con la de entrenamiento y la mejor época coincidente para pérdida y precisión de validación (época 19) indican que el modelo generaliza bien y no solo memoriza los datos de entrenamiento.

En el gráfico actual, hay una mejor alineación entre la pérdida de entrenamiento y

validación y menos volatilidad en la precisión de validación, lo que indica un modelo más estable.

Estadísticas de eficacia

| | | Confusion Matrix | | | | | | |
|--------|-------|------------------|-----|-----|----|-----|-----|------|
| Actual | AKIEC | 47 | 1 | 0 | 0 | 1 | 0 | 0 |
| | BCC | 0 | 77 | 0 | 0 | 0 | 0 | 0 |
| | BKL | 6 | 2 | 137 | 1 | 12 | 7 | 0 |
| | DF | 0 | 0 | 0 | 17 | 0 | 0 | 0 |
| | MEL | 2 | 0 | 8 | 1 | 142 | 14 | 0 |
| | NV | 2 | 15 | 44 | 5 | 69 | 862 | 9 |
| | VASC | 0 | 0 | 0 | 0 | 0 | 0 | 22 |
| | | AKIEC | BCC | BKL | DF | MEL | NV | VASC |
| | | Predicted | | | | | | |

Figura 4.8: Estadísticas de eficacia del modelo al estimar los resultados en el conjunto de pruebas

NV muestra la mayoría de los errores, lo que puede deberse a una mayor prevalencia en el conjunto de datos.

Comparando ambos gráficos, podemos identificar áreas específicas donde el modelo necesita mejora. Por ejemplo, la clase 'NV', a pesar de tener muchas predicciones correctas, tiene un número relativamente alto de falsos positivos y falsos negativos, lo que sugiere que puede haber características de las imágenes 'NV' que se confunden con otras clases. En cambio, 'BCC' y 'DF' parecen estar bien diferenciados del

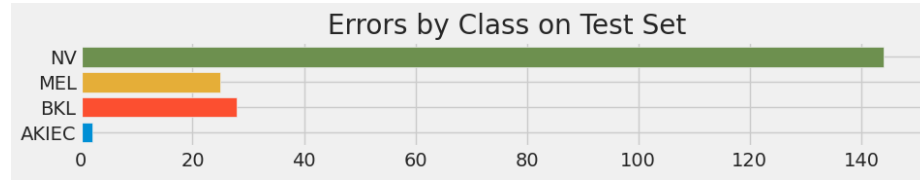


Figura 4.9: Gráfico de errores por clase en el conjunto de pruebas

resto. Esto puede informar estrategias futuras para mejorar la precisión del modelo, como la recolección de más datos o la implementación de técnicas de aprendizaje más avanzadas para clases específicas.

Informe de clasificación

Tabla 4.3: Informe de clasificación para el Experimento 2

| Categoría Diagnóstica | Precisión | Recall | F1-Score | Soporte |
|-----------------------|-----------|--------|----------|---------|
| AKIEC | 0.82 | 0.96 | 0.89 | 49 |
| BCC | 0.81 | 1.00 | 0.90 | 77 |
| BKL | 0.72 | 0.83 | 0.77 | 165 |
| DF | 0.71 | 1.00 | 0.83 | 17 |
| MEL | 0.63 | 0.85 | 0.73 | 167 |
| NV | 0.98 | 0.86 | 0.91 | 1006 |
| VASC | 0.71 | 1.00 | 0.83 | 22 |
| Accuracy | | | 0.87 | 1503 |
| Macro Avg | 0.77 | 0.93 | 0.84 | 1503 |
| Weighted Avg | 0.89 | 0.87 | 0.87 | 1503 |

Resultados Generales: La precisión global del modelo fue del 87%, con una media ponderada de precisión del 89%, recall del 87%, y una puntuación F1 del 87%. Resultados por Categoría Diagnóstica: Enumerar la precisión, el recall, y la puntuación F1 para cada categoría diagnóstica. Resalta que todas las categorías mejoraron en comparación con el Experimento 1, con 'NV' mostrando nuevamente la precisión más alta (0.98) y 'MEL' la más baja (0.63).

4.3. Comparación entre los Experimentos

Eficiencia del Modelo: El Experimento 2 demuestra una eficiencia general superior con una mayor precisión de validación y un modelo más estable. Esto puede atribuirse

a un balance más efectivo de clases y a ajustes en la tasa de aprendizaje.

Manejo de Desequilibrio de Clases: El Experimento 2 maneja mejor el desequilibrio de clases, lo que se refleja en una mejor diferenciación entre ciertas categorías diagnósticas.

Generalización: El Experimento 2 muestra una tendencia más estable en la pérdida de validación y en la precisión, sugiriendo una mejor capacidad para generalizar a nuevos datos.

Implicaciones para la Clasificación de Cáncer de Piel: Estos resultados resaltan la importancia de la estratificación de datos y del balance de clases en la clasificación de imágenes médicas, especialmente en contextos donde la variabilidad entre categorías es significativa.

4.3.1. Comparación de Ventajas y Desventajas

Ventajas del Experimento 1

1. **Sensibilidad:** El modelo del Experimento 1 demostró una alta sensibilidad, con un recall perfecto en varias categorías, lo que es crucial en aplicaciones médicas para evitar falsos negativos.
2. **Eficiencia en Categorías Específicas:** Ciertas categorías como 'NV' tuvieron una precisión extremadamente alta, lo que sugiere una eficiencia notable en la identificación de esta condición común.

Desventajas del Experimento 1

1. **Balance de Precisión-Recall:** Aunque el modelo tenía un alto recall en algunas categorías, la precisión era baja, lo que podría resultar en un número más alto de falsos positivos.
2. **Consistencia en el Rendimiento:** La variabilidad en el rendimiento entre categorías puede indicar una necesidad de ajustar el modelo o los datos para mejorar la consistencia.

Ventajas del Experimento 2

1. **Mejora en Precisión y F1-Score:** El modelo del Experimento 2 mostró mejoras en la precisión y puntuación F1 en todas las categorías, indicando un equilibrio más sólido entre la precisión y el recall.
2. **Rendimiento General:** Con una precisión global y una puntuación F1 más altas que el Experimento 1, este modelo demostró ser más robusto y equilibrado.

Desventajas del Experimento 2

1. Potencial Sobreajuste: A pesar de la mejora en los indicadores, un rendimiento demasiado alto en el conjunto de entrenamiento podría sugerir un sobreajuste, especialmente si no se replica en datos independientes.

4.4. Consideraciones Finales

Los resultados obtenidos son prometedores y sugieren que los modelos de aprendizaje profundo tienen un potencial considerable para mejorar la precisión y la eficiencia del diagnóstico del cáncer de piel

Conclusiones

Los resultados obtenidos muestran una alta eficacia (79% aproximadamente) que puede estar sujeta a futuras mejoras del modelo y del algoritmo. En general, los modelos de detección de cáncer de piel pueden tener una precisión del 85% al 95% o superior, dependiendo de la complejidad del problema, la calidad de los datos y la selección del algoritmo utilizado. Sin embargo, el 78% es un porcentaje elevado para los algoritmos de aprendizaje automático, y con algunas mejoras en el modelo puede incrementarse. También se desprende de este resultado la intención de convertir este proyecto en un producto a gran escala para ser utilizado, inicialmente, por los profesionales sanitarios de atención primaria, como prueba complementaria de alta efectividad a la hora de derivar a un paciente con una lesión cutánea al área de oncología. Con la premisa de convertir los resultados obtenidos en un producto al servicio de la sociedad, el equipo investigador pretende seguir investigando y desarrollando una solución mejor.

Recomendaciones

Datos: procesamiento y estrategias de normalización

1. Enriquecimiento de Datos: Complementar el conjunto de datos con imágenes adicionales de fuentes confiables para mejorar la robustez del modelo.
2. Aumentar la Diversidad de Datos: Utilizar técnicas de aumento de datos para generar variantes adicionales de las imágenes existentes, como rotaciones, zoom, o cambios en el brillo, para mejorar la capacidad del modelo de generalizar a partir de nuevos datos.
3. Balance de Clases mediante Aumento: Para categorías con menos ejemplos, aplicar técnicas de aumento de datos de manera selectiva para equilibrar la distribución de clases y reducir el sesgo del modelo.
4. Estratificación Mejorada: Asegurarse de que la estratificación de datos se aplique de manera que todas las particiones (entrenamiento, validación, prueba) tengan una distribución representativa de cada clase.
5. Conjuntos de Validación y Prueba Independientes: Considerar el uso de conjuntos de validación y prueba completamente independientes para obtener una evaluación más realista del rendimiento del modelo.

Modelo: optimización y ajuste de hiperparámetros

1. Selección de Características: Investigar técnicas de selección de características para identificar y utilizar solo los aspectos más relevantes de los datos de imagen para la clasificación.
2. Regularización y Arquitectura de Red: Explorar diferentes métodos de regularización, como dropout o penalizaciones L1/L2, y arquitecturas de red para encontrar el mejor equilibrio entre el rendimiento y la prevención del sobreajuste.

3. Tuning de Hiperparámetros: Realizar una búsqueda en cuadrícula o aleatoria más extensa para optimizar hiperparámetros como la tasa de aprendizaje, el tamaño del lote y el número de epochs.
4. Learning Rate Scheduling: Implementar un programa de ajuste de tasa de aprendizaje que disminuya la tasa basándose en el rendimiento del modelo durante el entrenamiento para mejorar la convergencia.
5. Análisis de Errores: Realizar un análisis de errores en profundidad para entender las causas de las clasificaciones incorrectas y abordarlas específicamente.
6. Presición y sesgos: Implementar modelos más complejos de clasificación para las clases menos representadas.

Bibliografía

- [1] '*Machine Learning*': definición, tipos y aplicaciones prácticas. Accessed: 2023-11-19. Iberdrola, 2023. URL: <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico> (vid. pág. V).
- [2] *Adamax*. <https://keras.io/api/optimizers/adamax/>. 2023 (vid. pág. 23).
- [3] Adekanmi Adegun y Serestina Viriri. «Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art». En: *Artificial Intelligence Review* 54 (2021), págs. 811-841 (vid. págs. 2, 10).
- [4] Departamento de Ingeniería Eléctrica. Universidad Nacional de Colombia Alberto Delgado PhD. Profesor Asociado. *Aplicación de las Redes Neuronales en Medicina*. Online; accessed 22-October-2023. 1999. URL: <https://repositorio.unal.edu.co/bitstream/handle/unal/32711/19460-64062-1-PB.pdf> (vid. págs. 2, 6-8, 10).
- [5] A Ameri. «A deep learning approach to skin cancer detection in dermoscopy images». En: *Journal of biomedical physics & engineering* 10.6 (2020), pág. 801 (vid. págs. 2, 10).
- [6] *Aumentación de datos*. <https://machinelearningmastery.com/image-augmentation-deep-learning-keras/>. 2023 (vid. pág. 20).
- [7] Kaustav Bera y col. «Predicting cancer outcomes with radiomics and artificial intelligence in radiology». En: *Nature Reviews Clinical Oncology* 19.2 (2022), págs. 132-146 (vid. pág. 8).
- [8] Titus Josef Brinker y col. «Skin cancer classification using convolutional neural networks: systematic review». En: *Journal of medical Internet research* 20.10 (2018), e11936 (vid. págs. 2, 10).
- [9] Lei Cai, Jingyang Gao y Di Zhao. «A review of the application of deep learning in medical image classification and segmentation». En: *Annals of translational medicine* 8.11 (2020) (vid. pág. 1).

- [10] Michael Calonder y col. «Brief: Binary robust independent elementary features». En: *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part IV 11*. Springer. 2010, págs. 778–792 (vid. pág. 7).
- [11] *Capas densas*. <https://towardsdatascience.com/activated-dense-layer-intuition-and-implementation-93e090cad34>. 2023 (vid. pág. 22).
- [12] Akshay S Chaudhari y col. «Low-count whole-body PET with deep learning in a multicenter and externally validated study». En: *NPJ digital medicine* 4.1 (2021), pág. 127 (vid. pág. 8).
- [13] *Color Histograms in Image Retrieval*. Disponible en: <https://www.pinecone.io/learn/series/image-search/color-histograms/>. 2023. URL: <https://www.pinecone.io/learn/series/image-search/color-histograms/> (vid. pág. 7).
- [14] *Cómo Extraer Features de una Imagen con HOG en Scikit-Image*. Último acceso: 2023-11-27. DataSmarts Español, 2020. URL: <https://datasmarts.net/es/como-extraer-features-de-una-imagen-con-hog-en-scikit-image/> (vid. pág. 7).
- [15] *Conjunto de datos HAM10000*. <https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000>. 2023 (vid. págs. 16, 19).
- [16] Kinnor Das y col. «Machine learning and its application in skin cancer». En: *International Journal of Environmental Research and Public Health* 18.24 (2021), pág. 13409 (vid. págs. 2, 3, 9).
- [17] *Dataframes de Pandas*. https://pandas.pydata.org/docs/getting_started/intro_tutorials/01_table_oriented.html. 2023 (vid. pág. 20).
- [18] Datamount. *What Canny Edge Detection Algorithm is All About*. Último acceso: 2023-12-01. 2023. URL: <https://medium.com/@datamount/what-canny-edge-detection-algorithm-is-all-about-103d94553d21> (vid. pág. 7).
- [19] Li Deng, Dong Yu y col. «Deep learning: methods and applications». En: *Foundations and trends® in signal processing* 7.3–4 (2014), págs. 197–387 (vid. pág. 1).
- [20] Fundación Piel Sana - Academia Española de Dermatología y Venereología. *¿Qué es la Dermatología?* <https://aedv.fundacionpielsana.es/piel-sana/que-es-la-dermatologia/>. Último acceso: 19 de noviembre de 2023 (vid. pág. 1).
- [21] Mehwish Dildar y col. «Skin cancer detection: a review using deep learning techniques». En: *International journal of environmental research and public health* 18.10 (2021), pág. 5479 (vid. pág. 11).

- [22] *Dropout*. <https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/>. 2023 (vid. pág. 22).
- [23] *EfficientNetB1*. <https://towardsdatascience.com/complete-architectural-details-of-efficientnets-7e50bc115fe0> (vid. pág. 21).
- [24] *Estadísticas importantes sobre el cáncer de piel tipo melanoma*. Último acceso: 2023-11-27. American Cancer Society, 2023. URL: <https://www.cancer.org/es/cancer/tipos/cancer-de-piel-tipo-melanoma/acerca/estadisticas-clave.html> (vid. págs. 1, 2).
- [25] L. Gerhardt. «Pattern recognition and machine learning». En: *IEEE Transactions on Automatic Control* 19.4 (1974), págs. 461-462. DOI: 10.1109/TAC.1974.1100578 (vid. pág. 6).
- [26] X. Du-Harpur y col. «What is AI? Applications of artificial intelligence to dermatology». En: *Frontiers in Medicine* 7 (2020). DOI: 10.3389/fmed.2020.00100. URL: <https://www.frontiersin.org/articles/10.3389/fmed.2020.00100/full> (vid. pág. 5).
- [27] *ImageDataGenerator*. <https://keras.io/api/preprocessing/image/>. 2023 (vid. pág. 20).
- [28] Silvia Jiménez de las Heras. *Clasificación de imágenes médicas mediante redes convolucionales*. Online; accessed 22-October-2023. 2018. URL: <http://hdl.handle.net/10486/688150> (vid. págs. 12, 13).
- [29] Yann LeCun, Yoshua Bengio y Geoffrey Hinton. «Deep learning». En: *nature* 521.7553 (2015), págs. 436-444 (vid. pág. 1).
- [30] Kang-Woo Lee y col. «Analysis of facial ultrasonography images based on deep learning». En: *Scientific reports* 12.1 (2022), pág. 16480 (vid. pág. 8).
- [31] Isidoro Gil Leiva, Pedro Díaz Ortuño y José Vicente Rodríguez Muñoz. *Técnicas y usos en la clasificación automática de imágenes*. 2019 (vid. pág. 7).
- [32] Tony Lindeberg. «Scale invariant feature transform». En: (2012) (vid. pág. 7).
- [33] Claudia Olavarrieta Martínez. «Ensemble de redes convolucionales para la clasificación de lesiones de cáncer de piel». Tesis de Licenciatura. Habana, Cuba: Universidad de la Habana, 2022 (vid. pág. 3).
- [34] *One-hot encoding*. <https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/>. 2023 (vid. pág. 19).
- [35] D1 Piccolo y col. «Concordance between telepathologic diagnosis and conventional histopathologic diagnosis: a multiobserver store-and-forward study on 20 skin specimens». En: *Archives of dermatology* 138.1 (2002), págs. 53-58 (vid. pág. 5).

- [36] Qindel. *Reconocimiento de imágenes: Qué es y cómo funciona*. Último acceso: 2023-12-01. 2023. URL: <https://www.qindel.com/que-es-y-como-funciona-el-reconocimiento-de-imagenes/> (vid. pág. 6).
- [37] *Regularización*. <https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-neural-networks-with-weight-regularization-in-keras/>. 2023 (vid. pág. 22).
- [38] G Romero y col. «Practice models in teledermatology in Spain: longitudinal study, 2009-2014». En: *Actas Dermo-Sifiliográficas (English Edition)* 109.7 (2018), págs. 624-630 (vid. pág. 5).
- [39] Weibin Rong y col. «An improved CANNY edge detection algorithm». En: *2014 IEEE international conference on mechatronics and automation*. IEEE. 2014, págs. 577-582 (vid. pág. 7).
- [40] Bhuvaneshwari Shetty y col. «Skin lesion classification of dermoscopic images using machine learning and convolutional neural network». En: *Scientific Reports* 12.1 (2022), pág. 18134 (vid. págs. 2, 10).
- [41] Vijay Singh. *Keras - Categorical Cross Entropy Loss Function*. Accessed: [25 de noviembre de 2023]. 2021. URL: <https://vitalflux.com/keras-categorical-cross-entropy-loss-function/> (vid. págs. 20, 23).
- [42] Hyun Ah Song y Soo-Young Lee. «Hierarchical representation using NMF». En: *Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part I 20*. Springer. 2013, págs. 466-473 (vid. pág. 1).
- [43] Amin Tajerian y col. «Design and validation of a new machine-learning-based diagnostic tool for the differentiation of dermatoscopic skin cancer images». En: *Plos one* 18.4 (2023), e0284437 (vid. págs. 2, 10).
- [44] Mingxing Tan y Quoc Le. *EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling*. Accessed: [24 de noviembre de 2023]. 2019. URL: <https://blog.research.google/2019/05/efficientnet-improving-accuracy-and.html> (vid. pág. 14).
- [45] Mingxing Tan y Quoc V. Le. «EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks». En: *ar5iv* (2019). URL: <https://ar5iv.labs.arxiv.org/html/1905.11946> (vid. pág. 15).
- [46] *Tasa de aprendizaje*. <https://www.jeremyjordan.me/nn-learning-rate/>. 2023 (vid. pág. 23).
- [47] Philipp Tschandl, Cliff Rosendahl y Harald Kittler. «The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions». En: *Scientific data* 5.1 (2018), págs. 1-9 (vid. pág. 19).

- [48] Analytics Vidhya. *How to Improve Class Imbalance using Class Weights in ML?* Online; accessed 16-November-2023. Oct. de 2020. URL: <https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/> (vid. pág. 24).
- [49] Darien Viera Barredo. «Autómata celular estocástico en redes complejas para el estudio de la invasión, migración y metástasis del cancer». Tesis de Licenciatura. Habana, Cuba: Universidad de la Habana, 2019 (vid. pág. 3).
- [50] John D Whited y col. «Teledermatology's impact on time to intervention among referrals to a dermatology consult service». En: *Telemedicine Journal and e-Health* 8.3 (2002), págs. 313-321 (vid. pág. 5).