

Universidad de La Habana
Facultad de Matemática y Computación



Título de la tesis

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

Dedicación

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. Mediante la implementación de modelos avanzados, como EfficientNet, y el ajuste de la tasa de aprendizaje, se pueden entrenar estos sistemas para clasificar distintos tipos de enfermedades de la piel, en especial el cáncer de piel, con alta precisión. Este trabajo de diploma presenta una serie de implementaciones, que incluyen CNN, el modelo EfficientNetB5 y el uso de Learning Rate Adjustment al dataset HAM10000 de imágenes de cáncer de piel, para lograr una detección efectiva. Estos resultados pueden sentar las bases para futuras investigaciones y mejoras en el campo de la detección temprana del cáncer de piel utilizando técnicas de aprendizaje automático y redes neuronales convolucionales.

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 [1]. Convolutional Neural Networks (CNNs), a specialization of this discipline, known for their efficiency in image processing, are ideal for detecting common features in dermatological images. By implementing advanced models, such as EfficientNet, and adjusting the learning rate, these systems can be trained to classify different types of skin diseases, especially skin cancer, with high accuracy. This diploma work presents a series of implementations, including CNN, the EfficientNetB5 model, and the use of Learning Rate Adjustment to the HAM10000 dataset of skin cancer images, to achieve effective detection. These results can lay the foundation for future research and improvements in the field of early skin cancer detection using machine learning techniques and convolutional neural networks.

Índice general

| | |
|--|-----------|
| Introducción | 1 |
| 1. Estado del Arte | 3 |
| 1.1. Teledermatología | 3 |
| 1.2. Técnicas tempranas de clasificación de imágenes | 4 |
| 1.3. Introducción y Desarrollo de Redes Neuronales | 5 |
| 1.3.1. Análisis de imágenes médicas | 5 |
| 1.3.2. Clasificación y reconocimiento de patrones en imágenes médicas | 6 |
| 1.4. Advenimiento de las redes neuronales convolucionales (CNN) | 7 |
| 1.5. Técnicas modernas y avances recientes | 9 |
| 1.6. Retos y limitaciones actuales | 9 |
| 1.6.1. Dificultades en la interpretación de resultados y la falta de explicabilidad de los modelos | 9 |
| 1.6.2. Problemas relacionados con el sobreajuste y la generalización de los modelos | 10 |
| 1.6.3. Limitaciones de los métodos tradicionales de aprendizaje automático en diagnóstico clínico | 10 |
| 1.6.4. Escasez de datos y variedad limitada en categorías de cáncer de piel | 10 |
| 1.7. Conclusión del estado del arte | 11 |
| 2. Propuesta | 12 |
| 2.1. Metodología | 12 |
| 2.1.1. Selección del modelo y transferencia de aprendizaje | 12 |
| 2.1.2. Preprocesamiento de datos y balanceo de clases | 12 |
| 2.1.3. Arquitectura del Modelo y Regularización | 13 |
| 2.1.4. Ajuste dinámico del learning rate | 13 |
| 2.1.5. Diversidad en propuestas de soluciones | 13 |

| | |
|--|-----------|
| 3. Detalles de implementación y experimentos | 14 |
| 3.1. Dataset | 14 |
| 3.2. Preparación y carga de datos | 14 |
| 3.2.1. Transformación de datos | 15 |
| 3.2.2. Modelación y división del conjunto de datos | 15 |
| 3.3. Generadores de datos y preprocesamiento | 17 |
| 3.3.1. Experimento 1 | 17 |
| 3.3.2. Experimento 2 | 17 |
| 3.3.3. Aumento y división de datos | 18 |
| 3.4. Diseño y entrenamiento del modelo | 19 |
| 3.4.1. ImageNet | 19 |
| 3.4.2. Arquitectura del modelo EfficientNet | 20 |
| 3.4.3. EfficientNetB1 | 20 |
| 3.4.4. Capas adicionales y regularización | 21 |
| 3.4.5. Optimización | 22 |
| 3.4.6. Ajuste dinámico de la tasa de aprendizaje | 22 |
| 3.5. Resultados | 23 |
| 3.5.1. Experimento 1 | 24 |
| 3.5.2. Experimento 2 | 27 |
| 3.5.3. Estadísticas básicas | 27 |
| 3.5.4. Estadísticas de aprendizaje | 28 |
| 3.5.5. Estadísticas de eficacia | 28 |
| Conclusiones | 32 |
| Recomendaciones | 33 |
| Bibliografía | 34 |

Índice de figuras

| | |
|--|----|
| 3.1. Estadísticas del rendimiento de los modelos de EfficientNet | 20 |
| 3.2. Estadísticas básicas del modelo EfficientNetB1. | 24 |
| 3.3. Estadísticas de aprendizaje a lo largo del proceso de entrenamiento. . | 25 |
| 3.4. Estadísticas de eficacia del modelo al estimar los resultados en el conjunto de pruebas | 26 |
| 3.5. Gráfico de errores por clase en el conjunto de pruebas | 27 |
| 3.6. Estadísticas básicas de algunas iteraciones del modelo EfficientNetB5. | 27 |
| 3.7. Estadísticas de aprendizaje a lo largo del proceso de entrenamiento. . | 28 |
| 3.8. Estadísticas de aprendizaje a lo largo del proceso de entrenamiento. . | 29 |
| 3.9. Estadísticas de eficacia del modelo al estimar los resultados en el conjunto de pruebas | 30 |
| 3.10. Gráfico de errores por clase en el conjunto de pruebas | 31 |

Ejemplos de código

Introducción

La dermatología es una rama especializada de la medicina que se centra en el estudio, diagnóstico, tratamiento, y manejo de las enfermedades y trastornos de la piel, el órgano más extenso del cuerpo humano [15]. El cáncer de piel es el tipo de cáncer más frecuente en todo el mundo, y su detección precoz es crucial. En este sentido, la tecnología puede marcar una diferencia significativa, pues dicho proceso suele ser estrictamente humano, dependiente del oncólogo y la biopsia, prueba que en algunos países puede resultar muy cara. Sin embargo, la integración de la tecnología al análisis dermatológico del cáncer es desafiante debido a la variabilidad en la apariencia de estas lesiones. En la resolución de esta problemática, los algoritmos de Machine Learning, específicamente las redes neuronales convolucionales (CNN), han mostrado gran eficiencia.

El contexto histórico y social del uso del aprendizaje automático (AM) para la detección del cáncer de piel en imágenes dermatológicas es un relato de rápida evolución tecnológica y colaboración interdisciplinar. En las últimas décadas, los avances en potencia computacional y el advenimiento del deep learning (aprendizaje profundo) han permitido un progreso significativo en el análisis automatizado de imágenes. Los progresos en el reconocimiento de melanomas a partir de imágenes dermatoscópicas, han demostrado que los sistemas automatizados pueden lograr un rendimiento diagnóstico comparable al de los expertos humanos [17].

Se ha pasado del análisis básico de imágenes digitales a sofisticados algoritmos capaces de identificar patrones sutiles en las imágenes de lesiones cutáneas. Este progreso se ha visto impulsado por la necesidad de mejorar el acceso a la atención dermatológica y la eficacia del cribado del cáncer de piel, especialmente en regiones con escasez de estos servicios.

Además, la integración de la analítica de big data con el crowdsourcing y la tecnología ML, como el proyecto DataDerm, ejemplifica los esfuerzos de colaboración para refinar y mejorar los algoritmos de diagnóstico a través de grandes conjuntos de datos de imágenes de la piel. Este enfoque colaborativo ha sido fundamental a la hora de entrenar y validar modelos de ML para garantizar que sean sólidos y fiables para su uso clínico.

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, pero la subjetividad y la variabilidad en la interpretación de las imágenes dermatoscópicas pueden llevar a diagnósticos erróneos o tardíos.

add citations Varios estudios han comparado el desempeño de algoritmos de aprendizaje profundo para la detección del cáncer de piel con análisis de especialistas del campo. Codella et al. y Haenssle et al. utilizaron conjuntos de datos extensos para entrenar algoritmos como InceptionV4 y demostraron que estos superaban a los dermatólogos en precisión y especificidad. Brinker et al. y Tschandl et al. también mostraron resultados similares, donde redes neuronales convolucionales (CNN) como ResNet50 eran más eficientes, especialmente en diagnósticos de melanoma. Estos estudios sugieren un papel prometedor para el ML en la asistencia a dermatólogos para un diagnóstico más preciso de melanoma [10].

Este trabajo se enfoca en mejorar la precisión y eficiencia en la detección del cáncer de piel utilizando Machine Learning. Se centra en el diagnóstico mediante imágenes dermatológicas obtenidas del dataset antes mencionado para desarrollar y validar un modelo basado en EfficientNetB5 para clasificar lesiones cutáneas. Las enfermedades de la piel utilizadas en el entrenamiento del modelo son: Melanoma (MEL), Nevus Melanocítico (NV), Carcinoma de Células Basales (BCC), Queratosis Actínica, Enfermedad de Bowen (carcinoma intraepitelial) (AKIEC), Queratosis Benigna (BKL), Dermatofibroma (DF), Lesión Vascular (VASC) [32].

Se llevó a cabo una revisión exhaustiva para evaluar el impacto de las técnicas de deep learning (aprendizaje profundo), en la que se analizaron diversos resultados de investigación y se presentaron mediante herramientas, gráficos, tablas y marcos para mejor comprensión. Se resalta en el estado del arte la teledermatología y las técnicas de clasificación de imágenes, abarcando desde los métodos tempranos hasta el advenimiento y desarrollo de redes neuronales y técnicas modernas. Luego, en la propuesta de la investigación, se analizó la metodología de red convolucional con el algoritmo de EfficientNetB5 y el ajuste de aprendizaje, sobre el conjunto de datos de imágenes que corresponden al concurso HAM10000 (Human Against Machine), así como el proceso de preparación y diseño del modelo. Finalmente, se aclaran los detalles de implementación y experimentos, incluyendo los resultados obtenidos, ventajas y desventajas de los mismos, seguido de las conclusiones y recomendaciones basadas en los hallazgos del estudio.

Capítulo 1

Estado del Arte

El cáncer de piel es el más frecuente del mundo *annadir cita* . A diferencia de muchos otros tipos de cáncer que se desarrollan internamente, el cáncer de piel se forma externamente y suele ser visible, lo que lo convierte en el cáncer que se ve". Esta visibilidad subraya la importancia de los exámenes de la piel, tanto realizados por uno mismo como por dermatólogos. El diagnóstico y el tratamiento oportunos son primordiales. La mayoría de los casos de cáncer de piel son tratables si se detectan precozmente, lo que subraya la necesidad de identificar la enfermedad con prontitud. La detección precoz no sólo salva vidas, sino que también evita que la enfermedad se vuelva peligrosa, desfigurante o incluso mortal.

1.1. Teledermatología

Las imágenes dermatológicas se utilizan para detectar el cáncer de piel, mediante el análisis de las lesiones cutáneas y la pigmentación. Esta es una técnica utilizada para examinar las lesiones cutáneas mediante un instrumento manual denominado dermatoscopio, que amplía la piel hasta 10 veces. Con el avance de la tecnología, la atención sanitaria ha experimentado una transformación. Se ha producido un aumento de dispositivos de diagnóstico y se ha desplazado hacia el desarrollo de nuevas competencias en estadística y psicología de la toma de decisiones médicas.

La teledermatología comenzó a ganar terreno en la década de 2000 como una manera de proporcionar consulta dermatológica a distancia. Un trabajo pionero es el de Whited et al [34], que exploró la eficacia de la teledermatología en 2002. El estudio constató que la teledermatología reducía significativamente el tiempo de intervención de los pacientes, con una mediana de 5 días para las derivaciones teledermatológicas frente a 28 días para las derivaciones tradicionales. El estudio también constató que la teledermatología era especialmente eficaz para pacientes con afecciones urgentes o semi-urgentes. El artículo destaca las posibles ventajas de la teledermatología para

mejorar el acceso a la asistencia y reducir los tiempos de espera de los pacientes.

1.2. Técnicas tempranas de clasificación de imágenes

Antes del auge de las redes neuronales y del aprendizaje profundo, había varias técnicas tradicionales utilizadas para la clasificación de imágenes. Estas técnicas se basaban en la extracción manual de características y en métodos de aprendizaje automático clásicos. La clave de la clasificación tradicional de imágenes radica en la extracción efectiva de características. Estas características deberían ser invariantes a pequeñas transformaciones y variaciones en las imágenes. Principalmente se tenía en cuenta:

1. El Histograma de Gradientes Orientados (HOG): Esta técnica captura la distribución de direcciones de gradientes (orientaciones) en una imagen. *add reference*
2. Scale-Invariant Feature Transform (SIFT): Es utilizado para la detección y descripción de características locales en imágenes. El descriptor SIFT es invariante a cambios de escala, rotación y parcialmente invariante a cambios de punto de vista y afín. *add reference*
3. Binary Robust Independent Elementary Features (BRIEF): Es un descriptor de características que utiliza un conjunto de pruebas binarias para describir una región de interés en una imagen. *add reference*
4. Color Histograms: describen la distribución de colores en una imagen y son útiles para tareas en las que el color es un indicador importante. *add reference*

En 1974, se publicó un artículo en "IEEE Transactions on Automatic Control"[16] relacionado con el reconocimiento de patrones y el aprendizaje automático, lo que indica que ya en la década de 1970 se estaban explorando estas técnicas para el reconocimiento de patrones en datos.

Con el tiempo, los investigadores comenzaron a aplicar algoritmos de aprendizaje automático tradicionales como Máquinas de Soporte Vectorial (SVM) y Árboles de Decisión para clasificar imágenes [20] basándose en características extraídas manualmente. Esta ha logrado grandes avances en las últimas décadas en las siguientes tres áreas: (1) desarrollo y uso de algoritmos de clasificación avanzados, como algoritmos de clasificación de subpíxeles, por campo y basados en el conocimiento; (2) uso de múltiples funciones de teledetección, incluida información espectral, espacial, multi-temporal y multisensor; y (3) incorporación de datos auxiliares en los procedimientos

de clasificación, incluidos datos como topografía, suelo, carreteras y datos censales. La evaluación de la precisión es una parte integral de un procedimiento de clasificación de imágenes. La evaluación de la precisión basada en la matriz de errores es el enfoque más comúnmente empleado para evaluar la clasificación por píxel, mientras que los enfoques difusos están ganando atención para evaluar los resultados de la clasificación difusa. [21].

Otro algoritmo de machine learning utilizado en las primeras etapas de la clasificación de imágenes es el árbol de decisión. Los árboles de decisión son un tipo de algoritmo de aprendizaje supervisado que se utiliza para la clasificación y la regresión. En el contexto de la clasificación de imágenes, un árbol de decisión podría usarse para tomar decisiones basadas en ciertos criterios, como el color, la forma, el tamaño, etc., para clasificar una imagen en una categoría específica.

Con el tiempo, se introdujeron otros algoritmos de aprendizaje automático como los Clasificadores de Bosques Aleatorios, y Clasificadores de Bayes Ingenuos, que también se aplicaron a la clasificación de imágenes utilizando características manuales. A medida que se avanzaba, se desarrollaron técnicas más avanzadas como las Redes Neuronales Artificiales y, eventualmente, las Redes Neuronales Convolucionales (CNN), que mostraron una capacidad significativa para aprender características directamente de los datos, reduciendo así la necesidad de extracción manual de características

1.3. Introducción y Desarrollo de Redes Neuronales

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 como ultrasonido, 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]

Ultrasonido : La aplicación de redes neuronales en el ultrasonido ha mejorado significativamente la detección y caracterización de lesiones. Mediante el análisis automatizado de imágenes, se pueden identificar patrones sutiles que podrían pasarse por alto en la evaluación humana.

Resonancia Magnética (RM) : En el ámbito de la RM, las redes neuronales han permitido un avance en la segmentación de imágenes, mejorando la precisión en la identificación de estructuras anatómicas y patológicas. Esto es especialmente relevante en la neurología, donde se utilizan para detectar y monitorizar enfermedades como la esclerosis múltiple.

Medicina Nuclear : La aplicación de redes neuronales en medicina nuclear, como en la tomografía por emisión de positrones (PET), ha mejorado la interpretación de los estudios, permitiendo una detección más temprana y precisa de diversas patologías, incluyendo el cáncer.

Radiología : En radiología, las redes neuronales han permitido el desarrollo de sistemas de ayuda al diagnóstico, especialmente en la detección de patologías pulmonares y mamarias. Estos sistemas pueden identificar anomalías en las radiografías y mamografías con un nivel de precisión que se acerca al de los especialistas humanos.

1.3.2. Clasificación y reconocimiento de patrones en imágenes médicas

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.

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 *add citations*. 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.

Otros algoritmos, como el k-Nearest Neighbors (KNN), se basan en medir la diferencia entre los valores de píxeles o distancias para comparar la similitud entre imágenes. Este enfoque es simple, pero requiere una configuración óptima de los hiperparámetros para obtener buenos resultados, lo que hace que estos algoritmos sean dependientes del problema que se está abordando.

Sin embargo, enfrentaron una serie de limitaciones significativas que 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.

Estas limitaciones condujeron a la necesidad de innovación y evolución en el campo de las redes neuronales. Esta necesidad llevó al surgimiento de un enfoque revolucionario en la forma de las Redes Neuronales Convolucionales (CNN), que abordaron muchas de estas limitaciones de manera efectiva.

1.4. Advenimiento de las redes neuronales convolucionales (CNN)

Mediante el uso de redes neuronales convolucionales, para facilitar la identificación de características relevantes en las imágenes médicas que pudieran ser indicativas de alguna condición médica particular las redes neuronales convolucionales han probado ser muy eficientes. Estas aprovecharon estructuras específicas de datos, como imágenes, para capturar patrones con mayor precisión, redujeron la carga computacional y mejoraron la generalización y la interpretabilidad. Gracias a estas mejoras, las CNN han sido fundamentales en el avance de la visión por computadora y han demostrado ser herramientas poderosas en una amplia variedad de aplicaciones, desde reconocimiento de objetos hasta diagnóstico médico.

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]

Una revisión sistemática de 2018 destacó el uso de CNNs para clasificar lesiones cutáneas. En este estudio, se revisaron 13 artículos que utilizaban CNNs para la

clasificación de lesiones cutáneas, y se encontró que los métodos más comunes eran aquellos que usaban una CNN ya entrenada con otro gran conjunto de datos y luego optimizaban sus parámetros para la clasificación de lesiones cutáneas. Las CNNs mostraron un alto rendimiento como clasificadores de lesiones cutáneas, aunque se señaló la dificultad de comparar diferentes métodos de clasificación debido a la utilización de conjuntos de datos no públicos en algunos casos. [7]

Sin embargo, las técnicas de aprendizaje automático, incluidos los modelos de aprendizaje profundo, han surgido como herramientas prometedoras para el análisis de imágenes médicas, especialmente imágenes de lesiones cutáneas, para la detección del cáncer de piel melanoma. ^A Deep Learning Approach to Skin Cancer Detection in Dermoscopy Images"[5] es un proyecto que sugiere un modelo de aprendizaje profundo para identificar cáncer de piel usando imágenes de lesiones en la piel. En esta investigación, se utilizaron 3400 imágenes de la colección de fotos dermatoscópicas HAM10000, que incluyen lesiones tanto melanoma como no melanoma. Se armó una red neuronal convolucional profunda para separar las imágenes en benignas y malignas. Se aplicó una técnica de aprendizaje transferido usando AlexNet como el modelo ya entrenado. El modelo sugerido toma la imagen tal cual es como entrada y descubre por su cuenta características valiosas en la imagen para clasificarla. Así, se salta los procesos complicados de segmentación de lesiones y extracción de características.

Un notable estudio publicado el 28 de octubre de 2022 publicado en Nature "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network"[25] se utilizó un subconjunto del conjunto de datos HAM10000 para clasificar lesiones cutáneas de imágenes dermatoscópicas y demostró la eficacia del uso de aprendizaje automático y redes neuronales convolucionales (CNN) en la clasificación de lesiones cutáneas a partir de imágenes dermatoscópicas. La metodología propuesta mostró resultados prometedores a la hora de distinguir entre lesiones malignas y benignas. El estudio compara la precisión de clasificación de distintos algoritmos de aprendizaje automático y modelos CNN. Se concluye que los modelos CNN proporcionan una mayor precisión en comparación con otros algoritmos de aprendizaje automático. En el sistema propuesto, alcanzan una precisión del 95,18% con el modelo CNN.

Un artículo publicado en PubMed ^Aan efficient deep learning-based skin cancer classifier for an imbalanced dataset"[3] propuso un eficaz clasificador de cáncer de piel basado en el aprendizaje profundo empleando el conjunto de datos HAM10000. Se utilizaron varios modelos de aprendizaje profundo como AlexNet, InceptionV3 y RegNetY-320 para clasificar el cáncer de piel. El rendimiento del marco propuesto fue mejor que el de los métodos convencionales. La precisión, la puntuación F1 y el valor de la curva ROC obtenidos con el marco propuesto fueron del 91 %, el 88,1 % y el 0,95.

1.5. Técnicas modernas y avances recientes

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 [12]

Además, se introdujo un enfoque innovador de combinación de técnicas de aprendizaje automático y aprendizaje profundo para abordar el problema de la detección del cáncer de piel [31]. El modelo de aprendizaje profundo en este estudio empleó redes neuronales de última generación para extraer características de las imágenes, mientras que el modelo de aprendizaje automático procesó estas características de la imagen para clasificar las lesiones cutáneas.

En otro artículo publicado en PubMed "Design and validation of a new machine-learning-based diagnostic tool for the differentiation of dermatoscopic skin cancer images"[27] se 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, una colección de 10015 imágenes dermatoscópicas recogidas a lo largo de 20 años, 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. La puntuación F1 para queratosis actínica, carcinoma basocelular, queratosis benigna, dermatofibroma, melanoma y lesiones vasculares fue consecutivamente de 0,63, 0,72, 0,70, 0,54, 0,58 y 0,80.

1.6. Retos y limitaciones actuales

1.6.1. 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. [19]

1.6.2. 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. [19]

1.6.3. 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. [19]

1.6.4. 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. [19]

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

Para abordar el problema de clasificación de tumores de cáncer de piel utilizando el dataset HAM1000, se han implementado varias estructuras y técnicas de Machine Learning. Estas elecciones demostraron ser un enfoque equilibrado entre precisión, eficiencia y capacidad de generalización. A continuación, se describen las estructuras utilizadas y se proporcionan explicaciones sobre su elección.

2.1. Metodología

2.1.1. Selección del modelo y transferencia de aprendizaje

Se optó por EfficientNetB1, una Red Neuronal Convolutiva (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 [28]. 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.

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

3.1. Dataset

El dataset utilizado como medio de aprendizaje para este proyecto es HAM1000-segmentation-and-classification [9]. 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 [32] 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 % |

3.2. Preparación y carga de datos

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 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 [22]

3.2.1. 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 add citation* 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.

3.2.2. Modelación y división del conjunto de datos

Los datos se modelan a partir de un dataframe de Pandas [11]. 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.

Experimento 1

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 3.1: Experimento 1: Distribución de imágenes de cáncer de piel en Entrenamiento, Test, y Validación Datasets

| Diagnostic category | Training | Validation | Testing |
|---------------------|----------|------------|---------|
| AKIEC | 122 | 4 | 1 |
| BCC | 487 | 12 | 15 |
| BKL | 1045 | 31 | 23 |
| DF | 108 | 5 | 2 |
| MEL | 1036 | 38 | 39 |
| NV | 6389 | 154 | 162 |
| VASC | 137 | 1 | 4 |

Experimento 2

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.

Tabla 3.2: Experimento 2: Distribución de imágenes de cáncer de piel en los conjuntos de Entrenamiento, Test y Validación

| Diagnostic category | Training | Validation | Testing |
|---------------------|----------|------------|---------|
| AKIEC | 88 | 19 | 19 |
| BCC | 359 | 77 | 77 |
| BKL | 769 | 164 | 164 |
| DF | 80 | 17 | 17 |
| MEL | 779 | 166 | 166 |
| NV | 4693 | 1005 | 1005 |
| VASC | 99 | 21 | 21 |

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

3.3. 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.

3.3.1. Experimento 1

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 remuestreo 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.

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 |

Como se evidencia anteriormente, al separar la data en clases el conjunto se mantiene desbalanceado. Se hace necesario aplicar un método llamado *Class Weighting* [33]. 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:

3.3.2. Experimento 2

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.

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 |

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 3.5: 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 |

3.3.3. Aumento y división de datos

Para la carga y procesamiento de imágenes se utilizó la clase `ImageDataGenerator` de Keras [18]. 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 3.6: 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 |

3.4. Diseño y entrenamiento del modelo

Se utiliza la arquitectura de red neural denominada *EfficientNetB1* [14]. 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 *EfficientNetB5* 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.

3.4.1. ImageNet

ImageNet es un proyecto de base de datos visual extenso, diseñado principalmente para la investigación en reconocimiento visual de objetos. Contiene más de 14 millones de imágenes, anotadas manualmente para indicar qué objetos se muestran, y más de 20,000 categorías, cada una con cientos o miles de imágenes. Este proyecto ha sido fundamental en el avance de la visión por computadora y el aprendizaje profundo, proporcionando un recurso de datos inmenso y diverso para entrenar algoritmos de inteligencia artificial.

La relevancia de ImageNet para el aprendizaje profundo se puso de manifiesto

con el éxito de AlexNet en el desafío ImageNet 2012 [23], donde logró una tasa de error top-5 del 15.3%, notablemente inferior a la de sus competidores. Este hito marcó un punto de inflexión en el campo, demostrando la viabilidad y la potencia de las redes neuronales convolucionales (CNN) cuando se combinan con unidades de procesamiento gráfico (GPU) para el entrenamiento.

3.4.2. Arquitectura del modelo EfficientNet

La familia de arquitecturas EfficientNet, desarrollada por los autores en [28], 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 224x224, EfficientNet B7 cuenta con 66 millones de parámetros y acepta imágenes de 600x600. [29]

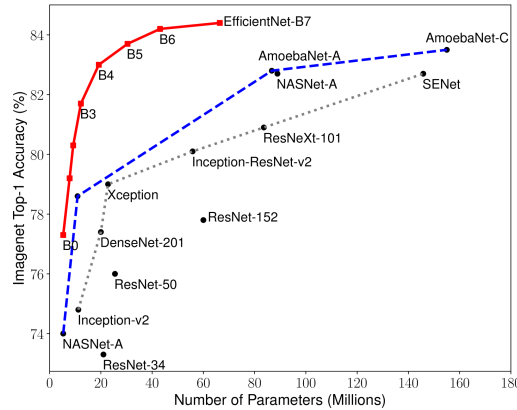


Figura 3.1: Estadísticas del rendimiento de los modelos de EfficientNet

3.4.3. EfficientNetB1

En relación con el dataset HAM1000 [9], 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.

3.4.4. 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. [[24]]

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. [8]

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. [13]

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.

3.4.5. 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* [26], idónea para problemas de clasificación multi-clase. Se configura el modelo para minimizarla y se rastrea la precisión como métrica principal.

3.4.6. 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 [30], 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.

3.5. 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 procesamiento y normalización de datos y optimizadores en el contexto del procesamiento de imágenes.

Tras la implementación del modelo, los resultados obtenidos superan las expectativas. 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:

Tabla 3.7: 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. |

3.5.1. Experimento 1

Estadísticas básicas

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

| 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 3.2: Estadísticas básicas del modelo EfficientNetB1.

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.

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 esta entrenando de forma correcta.

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

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

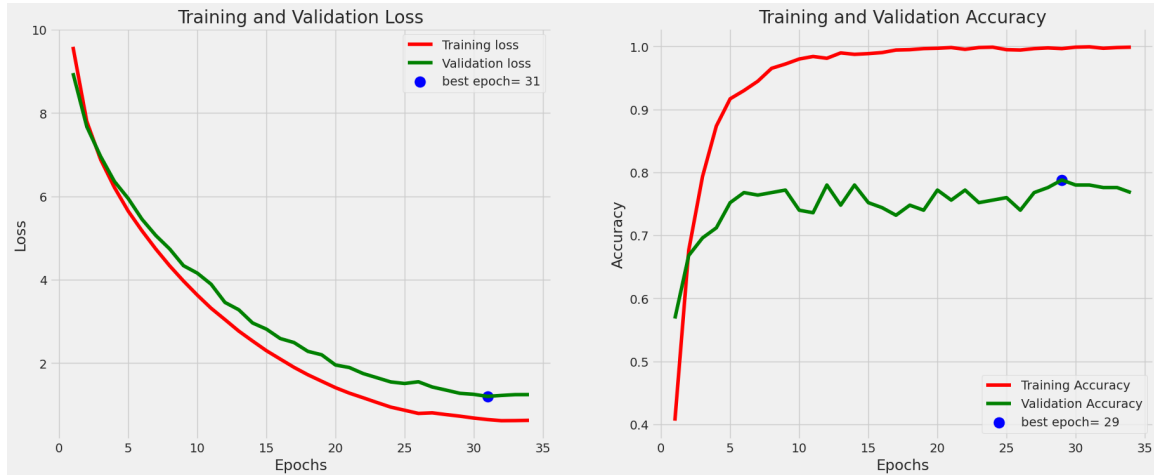


Figura 3.3: Estadísticas de aprendizaje a lo largo del proceso de entrenamiento.

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 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

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.

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.

| Confusion Matrix | | | | | | | | |
|------------------|-------|-----|-----|----|-----|----|------|---|
| Actual | AKIEC | 7 | 0 | 0 | 0 | 0 | 0 | |
| | BCC | 0 | 10 | 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 3.4: Estadísticas de eficacia del modelo al estimar los resultados en el conjunto de pruebas

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

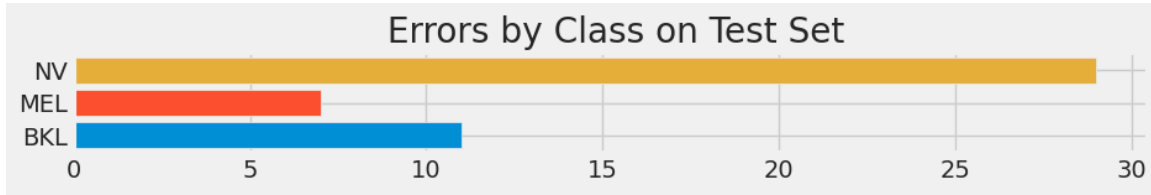


Figura 3.5: Gráfico de errores por clase en el conjunto de pruebas

3.5.2. Experimento 2

3.5.3. 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 3.6: Estadísticas básicas de algunas iteraciones del modelo EfficientNetB5.

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.

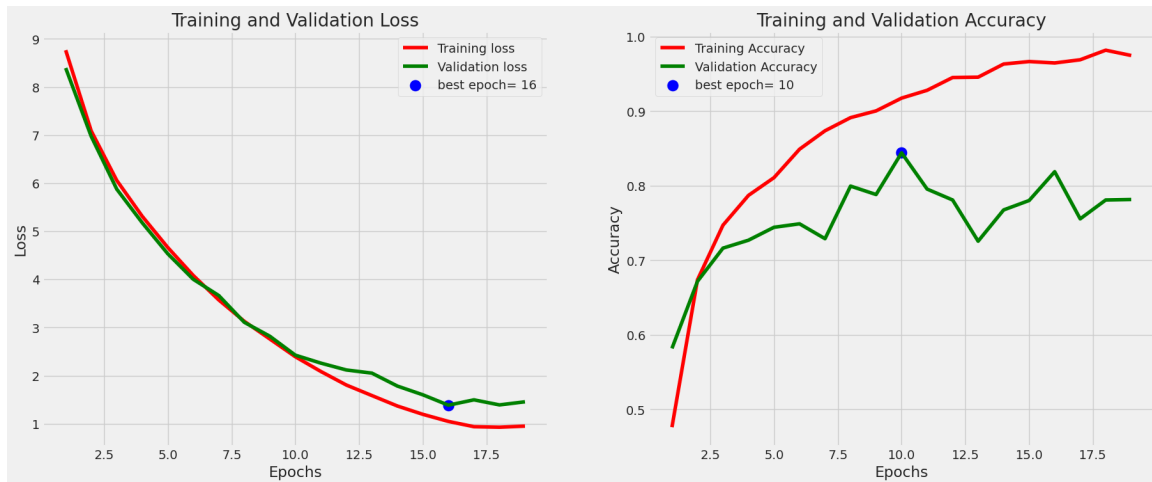


Figura 3.7: Estadísticas de aprendizaje a lo largo del proceso de entrenamiento.

3.5.4. 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 de aproximadamente 87 % (86.83%).

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.

3.5.5. Estadísticas de eficacia

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 mode-



Figura 3.8: Estadísticas de aprendizaje a lo largo del proceso de entrenamiento.

lo 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 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.

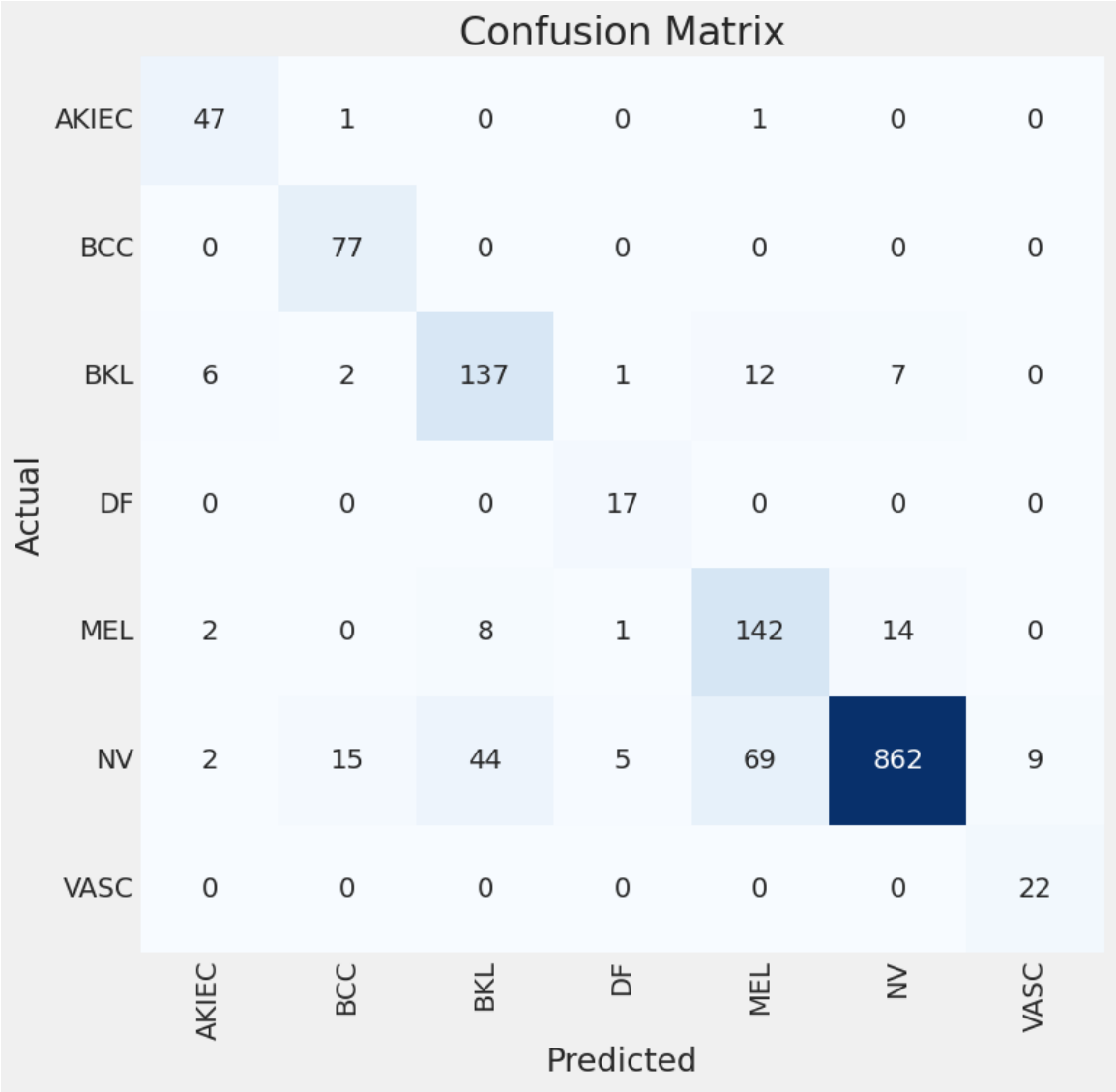


Figura 3.9: Estadísticas de eficacia del modelo al estimar los resultados en el conjunto de pruebas

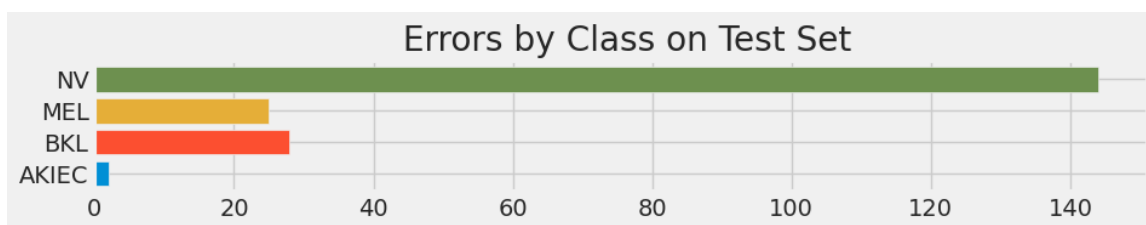


Figura 3.10: Gráfico de errores por clase en el conjunto de pruebas

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

Recomendaciones

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. págs. V, VI).
- [2] *Adamax*. <https://keras.io/api/optimizers/adamax/>. 2023 (vid. pág. 22).
- [3] Talha Mahboob Alam y col. «An efficient deep learning-based skin cancer classifier for an imbalanced dataset». En: *Diagnostics* 12.9 (2022), pág. 2115 (vid. pág. 8).
- [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. 5-7).
- [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. pág. 8).
- [6] *Aumentación de datos*. <https://machinelearningmastery.com/image-augmentation-deep-learning-keras/>. 2023 (vid. pág. 18).
- [7] 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. pág. 8).
- [8] *Capas densas*. <https://towardsdatascience.com/activated-dense-layer-intuition-and-implementation-93e090cad34>. 2023 (vid. pág. 21).
- [9] *Conjunto de datos HAM10000*. <https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000>. 2023 (vid. págs. 14, 20).
- [10] 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. pág. 2).

- [11] *Dataframes de Pandas*. https://pandas.pydata.org/docs/getting_started/intro_tutorials/01_table_oriented.html. 2023 (vid. pág. 15).
- [12] 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. 9).
- [13] *Dropout*. <https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/>. 2023 (vid. pág. 22).
- [14] *EfficientNetB1*. <https://towardsdatascience.com/complete-architectural-details-of-efficientnets-7e50bc115fe0> (vid. pág. 19).
- [15] 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).
- [16] 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. 4).
- [17] 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. 1).
- [18] *ImageDataGenerator*. <https://keras.io/api/preprocessing/image/>. 2023 (vid. pág. 18).
- [19] 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. pág. 10).
- [20] Òscar Lorente, Ian Riera y Aditya Rana. «Image classification with classic and deep learning techniques». En: *arXiv preprint arXiv:2105.04895* (2021) (vid. pág. 4).
- [21] Dengsheng Lu y Qihao Weng. «A survey of image classification methods and techniques for improving classification performance». En: *International journal of Remote sensing* 28.5 (2007), págs. 823-870 (vid. pág. 5).
- [22] *One-hot encoding*. <https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/>. 2023 (vid. pág. 15).
- [23] Pinecone. *ImageNet - The Dataset that Transformed Image Classification*. 2021. URL: <https://www.pinecone.io/learn/series/image-search/imagenet/> (vid. pág. 20).

- [24] *Regularización*. <https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-neural-networks-with-weight-regularization-in-keras/>. 2023 (vid. pág. 21).
- [25] 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. pág. 8).
- [26] 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. pág. 22).
- [27] 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. pág. 9).
- [28] 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. págs. 12, 20).
- [29] 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. 20).
- [30] *Tasa de aprendizaje*. <https://www.jeremyjordan.me/nn-learning-rate/>. 2023 (vid. pág. 22).
- [31] Jitendra V Tembhurne y col. «Skin cancer detection using ensemble of machine learning and deep learning techniques». En: *Multimedia Tools and Applications* (2023), págs. 1-24 (vid. pág. 9).
- [32] Philipp Tschandl, Cliff Rosendahl y Harald Kittler. «The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions». En: *Scientific data* 5.1 (2018), págs. 1-9 (vid. págs. 2, 14).
- [33] 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. 17).
- [34] 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. 3).