

Clasificación de imágenes indeterminadas de cáncer cutáneo a través de técnicas de análisis de datos

Leidy Tatiana Molano Muñoz

Asesor

Freddy Alexander Torres Payoma, PhD(c)

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería (ECBTI)

Especialización en Ciencia de Datos y Analítica

Semillero de Investigación Mathphysics

2024

Agradecimientos

Deseo expresar mi más sincero y profundo agradecimiento a todas las personas que, de una manera u otra, contribuyeron a la realización de este trabajo. Este logro no habría sido posible sin el apoyo, la guía y el aliento que recibí de quienes me rodearon durante este proceso.

A mi familia, Harold Molano, Socorro Muñoz, Daniel Ortega y Emma Lucia, quiero dedicarles estas líneas en reconocimiento a su amor incondicional, paciencia y respaldo constante. Su presencia en mi vida ha sido un pilar fundamental, dándome la fortaleza y la motivación necesarias para superar cada desafío y alcanzar esta meta. Gracias por ser mi refugio en los momentos difíciles y por celebrar conmigo cada pequeño avance. Este trabajo es tan suyo como mío.

A mi asesor, Freddy Alexander Torres Payoma, extiendo mi gratitud más sincera por su dedicación, profesionalismo y compromiso. Sus valiosas sugerencias, su paciencia infinita y su visión experta han sido determinantes para el desarrollo de este trabajo. Cada comentario, orientación y consejo me han permitido crecer no solo como profesional, sino también como persona. Agradezco profundamente el tiempo y esfuerzo que ha dedicado a este proyecto, convirtiéndose en un mentor ejemplar.

Asimismo, agradezco a la institución académica y a los docentes que a lo largo de mi formación profesional contribuyeron a construir los cimientos de este logro. Gracias por compartir su sabiduría y fomentar en mí una actitud crítica y proactiva frente al aprendizaje.

Resumen

El trabajo desarrolla un modelo basado en aprendizaje automático para clasificar imágenes dermatoscópicas indeterminadas de lesiones cutáneas, un desafío crucial en el diagnóstico temprano del cáncer de piel. Utilizando bases de datos públicas como ISIC (International Skin Imaging Collaboration), se aplicaron técnicas de preprocesamiento y aprendizaje por transferencia con modelos preentrenados como ResNet, mobilenet y Xception.

El modelo de redes neuronales convolucionales (CNN) alcanzó resultados competitivos en métricas como precisión, sensibilidad y AUC (Área bajo la curva), proporcionando además análisis de patrones útiles para los especialistas. Esta herramienta mejora la precisión diagnóstica, reduce la necesidad de biopsias invasivas y contribuye a la implementación de soluciones tecnológicas que optimizan la atención médica y la calidad de vida de los pacientes.

Palabras claves: Cáncer de piel, análisis de datos, imágenes indeterminadas, aprendizaje automático.

Abstract

The work develops a machine learning-based model to classify indeterminate dermoscopic images of skin lesions, a crucial challenge in the early diagnosis of skin cancer. Using public databases such as ISIC (International Skin Imaging Collaboration), preprocessing and transfer learning techniques were applied with pretrained models such as ResNet, mobilenet and Xception.

The convolutional neural network (CNN) model achieved competitive results in metrics such as accuracy, sensitivity and AUC (area under the curve), also providing useful pattern analysis for specialists. This tool improves diagnostic accuracy, reduces the need for invasive biopsies and contributes to the implementation of technological solutions that optimize medical care and patients' quality of life.

Keywords: Skin cancer, data analysis, indeterminate imaging, machine learning.

Tabla de Contenido

Problema de Investigación	12
Introducción	12
Antecedentes del Problema de Investigación	14
Estudios Referentes a Trabajos sobre Redes Neuronales Artificiales	15
Justificación	18
Objetivos	20
Objetivo General	20
Objetivos Específicos	20
Metodología	21
Fundamentación Teórica	24
La Piel y su Importancia	24
Cáncer de Piel	26
Tipos de Cáncer de Piel	26
Cáncer de Piel No Melanoma.	26
Melanoma Maligno.	27
Factores de Riesgo	27
Prevención y Diagnóstico	27
Tratamiento	28
Procesamiento de Imágenes Médicas	28
RGB: Modelo de Color	29
Aprendizaje Automático (Machine Learning)	30

Aprendizaje Supervisado	30
Aprendizaje No Supervisado	31
Aprendizaje Profundo (Deep Learning)	32
Redes Neuronales Artificiales	33
Estructura de una Red Neuronal Artificial	33
Capa de Entrada.	33
Capas Ocultas.	33
Capa de Salida.	34
Funcionamiento y Entrenamiento	34
Propagación Hacia Adelante.	34
Cálculo de Error.	34
Retropropagación.	34
Tipos de Redes Neuronales	34
Redes Neuronales Multicapa (MLP).	34
Redes Neuronales Convolucionales (CNN).	35
Redes Neuronales Recurrentes (RNN).	35
Redes Generativas (GAN).	35
Desarrollo del Modelo Basado en Deep Learning	36
Dataset	36
Comprensión de los Datos	37
Variedad de Lesiones.	37
Metadata Clínica.	37

Anotaciones Expertas.	37
Pre-Procesamiento de los Datos	37
Etapas del Pre-Procesamiento	38
Normalización.	38
Redimensionamiento.	38
Desarrollo del Modelo	39
Configuración de las Redes Pre-entrenadas	39
Parámetros de Entrenamiento para las Redes	40
Parámetros Clave de Configuración.	40
Ajuste de las Capas del Modelo.	41
Aislamiento de Capas Convolucionales.	41
Habilitación del Entrenamiento.	41
Adición de Capas Personalizadas.	43
Global Average Pooling.	43
Capa Densa de Salida.	43
Compilación del Modelo.	43
Optimizador Adam.	43
Función de Pérdida Categorical_Crossentropy.	43
Métricas de Evaluación.	43
Entrenamiento del Modelo.	44
Análisis de Resultados	46
MobileNet	46

Precisión y Validación Durante el Entrenamiento	46
Perdida Durante el Entrenamiento	47
Curva ROC y Curva AUC	48
Xception	49
Precisión y Validación Durante el Entrenamiento	50
Perdida Durante el Entrenamiento	51
Curva ROC y Curva AUC	52
ResNet101V2	53
Precisión y Validación Durante el Entrenamiento	54
Perdida Durante el Entrenamiento	54
Curva ROC y Curva AUC	56
Definiciones para el Calculo de las Gráficas	57
Barras de Error	57
Puntos de Datos	57
Lineas de Referencia	57
Desempeño de las Redes	57
Análisis de la Exactitud (ACC)	58
Análisis de la Sensibilidad (Sens)	60
Análisis de la Especificidad (Spec)	61
Análisis del Área Bajo la Curva (AUC)	62
Evaluación Estadística del Desempeño de los Modelos	64
Definiciones Estadísticas	64
Media.	64

Desviación Estándar.	65
Intervalo de Confianza.	65
Análisis Estadístico	65
Variabilidad	66
Precisión de las Estimaciones	67
Comparación Entre Redes	67
Conclusiones y Recomendaciones.....	69
Conclusiones	69
Recomendaciones	70
Referencias	72

Lista de Tablas

Tabla 1 <i>Resumen sobre Metodologías Aplicadas para la Detección del Cáncer de Piel</i> . . .	16
Tabla 2 <i>Resultados de la Clasificación de los Modelos</i>	58
Tabla 3 <i>Estadísticas Detalladas para Cada Métrica y Red</i>	66

Lista de Figuras

Figura 1	<i>Descripción de la Metodología</i>	21
Figura 2	<i>Descripción de las Capas de Piel</i>	25
Figura 3	<i>Representación de una Imagen en el Espacio RGB</i>	29
Figura 4	<i>Normalización de la Imagen</i>	38
Figura 5	<i>Red Convolutiva</i>	41
Figura 6	<i>Ilustración de la Configuración de las Capas de la Red</i>	42
Figura 7	<i>Ilustración de la Compilación del Modelo</i>	44
Figura 8	<i>Gráfica de Precisión y Validación para el Modelo MobileNet</i>	47
Figura 9	<i>Pérdida Durante el Entrenamiento para el Modelo MobileNet</i>	48
Figura 10	<i>Curva ROC (Receiver Operating Characteristic) para el Modelo MobileNet</i>	49
Figura 11	<i>Gráfica de Precisión y Validación para el Modelo Xception</i>	51
Figura 12	<i>Gráfica de Perdida Durante el Entrenamiento para el Modelo Xception</i>	52
Figura 13	<i>Curva ROC (Receiver Operating Characteristic) para el mModelo Xception</i>	53
Figura 14	<i>Gráfica de Precisión y Validación para el Modelo ResNet</i>	54
Figura 15	<i>Gráfica de Perdida para el Modelo ResNet</i>	55
Figura 16	<i>Curva ROC (Receiver Operating Characteristic) para el Modelo ResNet</i>	56
Figura 17	<i>Desempeño de la Exactitud de las Redes</i>	59
Figura 18	<i>Desempeño de la Sensibilidad de las Redes</i>	61
Figura 19	<i>Desempeño de la Especificidad de los Modelos</i>	62
Figura 20	<i>Desempeño del Área bajo la Curva de las Redes</i>	63

Problema de Investigación

Introducción

El cáncer de piel se posiciona como una de las principales preocupaciones de salud pública a nivel global, debido a su alta incidencia y la posibilidad de consecuencias graves si no se detecta y trata de manera oportuna. En Colombia, durante el 2023, uno de cada nueve nuevos casos de cáncer correspondió al tipo de cáncer de piel, ubicándose en el octavo puesto de los once tipos de cáncer más comunes según el Ministerio de Salud. Este panorama subraya la importancia de desarrollar estrategias eficaces para su diagnóstico y manejo temprano (Forero, 2024).

Entre los tipos de cáncer de piel más frecuentes se halla el carcinoma basocelular (CCB), el carcinoma escamocelular (CCE) y la queratosis actínica (QA), los cuales, aunque inicialmente pueden parecer menos agresivos, tienen el potencial de propagarse y causar metástasis rápidamente si no se diagnostican y tratan adecuadamente. Por otro lado, el melanoma maligno (MM), aunque menos frecuente, es el más letal. La urgencia de un manejo clínico oportuno para este tipo de cáncer es evidente, considerando que causa el 80 % de las muertes relacionadas con cáncer de piel (A. F. P. López, Álvarez, Caicedo, y Moreira, 2024).

El diagnóstico temprano del cáncer de piel requiere la realización de diversas pruebas, siendo el examen físico inicial fundamental para identificar lunares o manchas con características anormales. Una herramienta clave en este proceso es la dermatoscopia o microscopía de epiluminiscencia (ELM), una técnica no invasiva que permite obtener imágenes ampliadas de la lesión, entre 10 y 400 veces su tamaño original. Esto facilita al dermatólogo analizar patrones específicos en la imagen para orientar el diagnóstico. No obstante, cuando las características de la

lesión no son concluyentes, se procede a realizar una biopsia, un procedimiento invasivo que implica la extracción de una muestra de la lesión para su análisis posterior en el laboratorio, garantizando así un diagnóstico más preciso. A pesar de los avances tecnológicos, la tarea de categorizar correctamente las imágenes de lesiones cutáneas sigue siendo compleja y presenta un desafío significativo para los dermatólogos (Huix Pérez, 2024). Este reto se magnifica en escenarios con recursos limitados, donde el acceso a herramientas avanzadas es escaso.

En este contexto, el desarrollo de sistemas automatizados para la clasificación de imágenes de lesiones cutáneas ha adquirido una importancia creciente. La implementación de técnicas de análisis de datos, aprendizaje automático y procesamiento de imágenes promete mejorar significativamente la precisión diagnóstica y, en última instancia, reducir las tasas de mortalidad asociadas al cáncer cutáneo. Sin embargo, una de las principales dificultades radica en las imágenes indeterminadas, aquellas en las que las características visuales no permiten una clasificación clara entre lesiones benignas y malignas. Este tipo de imágenes representa una proporción considerable de los casos, lo que subraya la necesidad de soluciones innovadoras y eficientes.

El presente trabajo de grado tiene como objetivo principal desarrollar un modelo para la clasificación de imágenes indeterminadas de cáncer cutáneo utilizando técnicas avanzadas de análisis de datos. Este enfoque combina la integración de bases de datos públicas con algoritmos de aprendizaje profundo, considerando tanto las características visuales como los patrones estadísticos de las lesiones. Además, se busca evaluar el desempeño del modelo en escenarios clínicos simulados para medir su aplicabilidad en la práctica médica.

En el desarrollo de este estudio, se aborda no solo el diseño técnico del modelo, sino también la importancia de validar su impacto en la toma de decisiones clínicas. Con ello, se pretende

contribuir a la optimización del diagnóstico temprano y al tratamiento oportuno del cáncer cutáneo, sentando las bases para futuras investigaciones que integren inteligencia artificial y medicina.

Antecedentes del Problema de Investigación

En esta sección se contextualiza el presente estudio mediante una revisión detallada de los trabajos previos más relevantes en el campo de la clasificación de imágenes dermatológicas y las redes neuronales.

El cáncer de piel, en sus diversas manifestaciones, ha experimentado un incremento significativo en su incidencia a nivel mundial durante las últimas décadas. Según datos de la Organización Mundial de la Salud (OMS), el cáncer de piel no melanoma es el tipo de cáncer más común a nivel global, mientras que el melanoma maligno, aunque menos frecuente, es el más letal debido a su alta capacidad de generar metástasis (Ramírez y cols., 2024). En América Latina, y específicamente en Colombia, la creciente prevalencia de esta enfermedad ha llevado a los organismos de salud pública a priorizar su detección temprana como un factor clave para reducir la mortalidad asociada. El diagnóstico del cáncer de piel depende en gran medida de la experiencia del médico y del uso de herramientas como la dermatoscopia y la biopsia (Hernández, 2024). Sin embargo, uno de los principales retos radica en la clasificación de imágenes de lesiones cutáneas indeterminadas, que presentan características ambiguas y dificultan la distinción entre lesiones benignas y malignas. Este desafío se traduce en diagnósticos tardíos, tratamientos inadecuados y, en algunos casos, un aumento en la mortalidad del paciente.

En el ámbito de la investigación, se han realizado importantes avances en el uso de tecnologías emergentes para apoyar el diagnóstico médico. Los sistemas de análisis de imágenes y el aprendizaje automático han demostrado ser herramientas prometedoras para mejorar la precisión

diagnóstica en diversas áreas de la medicina. En particular, el desarrollo de modelos basados en redes neuronales convolucionales (CNN) ha permitido la automatización del análisis de imágenes cutáneas, logrando resultados comparables a los de dermatólogos expertos en algunos estudios recientes (Hernández, 2024). A pesar de estos avances, la implementación de estas tecnologías en contextos reales aún enfrenta limitaciones, como la variabilidad en la calidad de las imágenes, la diversidad de características étnicas y geográficas de los pacientes, y la falta de bases de datos suficientemente representativas. Además, el análisis de imágenes indeterminadas sigue siendo un área poco explorada, lo que resalta la necesidad de investigaciones enfocadas en este tipo de lesiones (Hernández, 2024). En Colombia, donde las tasas de cáncer de piel han ido en aumento, existen escasos estudios locales que aborden el uso de técnicas avanzadas de análisis de datos para mejorar el diagnóstico y la clasificación de lesiones cutáneas. Esta brecha en el conocimiento subraya la pertinencia del presente trabajo de investigación, cuyo objetivo es desarrollar un modelo eficaz para la clasificación de imágenes indeterminadas de cáncer cutáneo mediante técnicas de análisis de datos, con el fin de contribuir al diagnóstico temprano y la optimización de los tratamientos en el ámbito clínico. Teniendo en cuenta lo mencionado anteriormente, surge la pregunta de investigación:

¿Es posible clasificar las imágenes indeterminadas de cáncer cutáneo mediante el uso de técnicas avanzadas de análisis de datos, para optimizar el diagnóstico temprano y reducir la incertidumbre en la toma de decisiones clínicas?

Estudios Referentes a Trabajos sobre Redes Neuronales Artificiales

Para el desarrollo de la investigación se citan en la Tabla 1 los estudios referentes a trabajos que abordaron el problema desde diferentes horizontes:

Tabla 1*Resumen sobre Metodologías Aplicadas para la Detección del Cáncer de Piel*

Autor	Metodología	Resumen del Contenido
Vargas D. et al. (2021)	Redes Neuronales Convolucionales (CNN)	Se emplean CNN para clasificar imágenes dermatoscópicas y mejorar el diagnóstico temprano del cáncer de piel, demostrando una alta precisión en la identificación de lesiones.
Martínez González (2021)	Aprendizaje por Transferencia	Implementa un modelo basado en aprendizaje por transferencia, utilizando datasets ISIC, PH2 y EDRA y herramientas de visión artificial como TensorFlow y Keras.
Jaime Calva et al. (2022)	Redes Neuronales Convolucionales (CNN)	Implementación de CNN para clasificar imágenes de cáncer de piel, mostrando resultados prometedores para la detección temprana.
Martín Tinaquero, D. (2023)	Aprendizaje Profundo y Transformadores	Desarrolla una herramienta que combina redes convolucionales y modelos basados en transformadores para apoyar en el diagnóstico temprano y autoexamen de lesiones cutáneas.

Continúa en la siguiente página

Continúa en la siguiente página

Autor	Metodología	Resumen del Contenido
Fernández Morales, E. (2024)	Inteligencia Artificial y Algoritmos de Aprendizaje Automático	Desarrollo de un modelo para clasificar imágenes médicas y reducir falsos positivos/negativos, integrando imágenes DICOM y datos reales del Hospital Virgen del Rocío.

Nota. La tabla presenta un resumen de diferentes investigaciones en detección de cáncer de piel.

Justificación

El cáncer de piel es una enfermedad de alta prevalencia y potencial letalidad, especialmente en su forma más agresiva, el melanoma maligno. La detección y el diagnóstico temprano son esenciales para reducir la mortalidad, pero esto se ve obstaculizado por la dificultad de clasificar imágenes indeterminadas, es decir, aquellas cuyas características visuales no permiten una clara distinción entre lesiones benignas y malignas. Esta incertidumbre diagnóstica puede retrasar el inicio del tratamiento, lo que incrementa el riesgo de progresión y metástasis. En el contexto colombiano, el cáncer de piel ha ganado relevancia como un problema de salud pública, ocupando un lugar destacado entre los tipos de cáncer priorizados por las autoridades sanitarias. Sin embargo, el acceso limitado a tecnologías avanzadas y la variabilidad en la calidad de los diagnósticos representan barreras significativas para su manejo adecuado. Estas limitaciones son especialmente evidentes en zonas rurales o con recursos médicos restringidos, donde los diagnósticos dependen en gran medida de la observación clínica, lo que puede dar lugar a errores o retrasos.

El avance de las tecnologías de análisis de datos, incluyendo técnicas de aprendizaje automático y procesamiento de imágenes, ofrece una oportunidad sin precedentes para abordar este desafío. Estas herramientas han demostrado su capacidad para mejorar la precisión diagnóstica al identificar patrones complejos en imágenes médicas que pueden pasar desapercibidos para el ojo humano. Sin embargo, su aplicación en el análisis de imágenes indeterminadas de cáncer cutáneo sigue siendo limitada, y la mayoría de los estudios existentes se han centrado en lesiones claramente definidas.

Este estudio propone un nuevo modelo de análisis de imágenes para clasificar lesiones cutáneas dudosas, con el objetivo de mejorar la precisión diagnóstica, reducir biopsias innecesarias

y optimizar recursos médicos. Asimismo, busca impulsar la innovación tecnológica en salud en Colombia y mejorar los resultados clínicos para los pacientes.

Objetivos

Objetivo General

Desarrollar un modelo basado en técnicas avanzadas de análisis de datos para la clasificación de imágenes indeterminadas de cáncer cutáneo, con el propósito de optimizar el diagnóstico temprano y apoyar la toma de decisiones clínicas en el manejo de esta enfermedad.

Objetivos Específicos

Recopilar del conjunto de datos ISIC (International Skin Imaging Collaboration), imágenes de lesiones cutáneas, incluyendo tanto imágenes indeterminadas como aquellas con diagnóstico confirmado.

Implementar técnicas de preprocesamiento de imágenes y utilizar métodos de machine learning y deep learning para entrenar un modelo de clasificación que mejore la calidad de los datos y la precisión en la clasificación de imágenes.

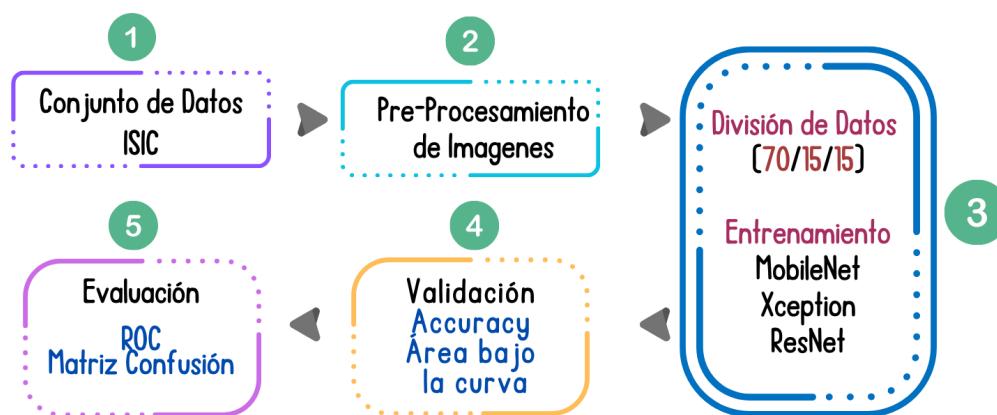
Analizar el rendimiento del modelo en términos de exactitud, precisión, sensibilidad, especificidad y AUC ROC.

Metodología

Para llevar a cabo el estudio se adoptó un enfoque metodológico que integra técnicas avanzadas de análisis de datos, aprendizaje automático y procesamiento de imágenes (Martínez-González, 2021; Torres Payoma, 2019). El proceso se dividió en cinco etapas principales: recopilación de datos, preprocesamiento, diseño e implementación del modelo, evaluación y validación, y análisis de resultados, tal como se observa en Figura 1.

Figura 1

Descripción de la Metodología



Nota. La figura muestra el proceso descriptivo para la implementación de la metodología. Autor, 2024.

En la primera etapa, recopilación de datos, se seleccionaron bases de datos públicas reconocidas, como ISIC, que contienen imágenes dermatoscópicas de lesiones cutáneas etiquetadas. Estas bases de datos fueron seleccionadas debido a su calidad, diversidad y representatividad en términos de diferentes tipos de cáncer de piel. En esta fase, también se priorizó la identificación de imágenes indeterminadas, aquellas cuya clasificación inicial es ambigua y representa un desafío diagnóstico.

La etapa de preprocesamiento de datos consistió en limpiar, normalizar y transformar las imágenes recopiladas. Para garantizar la homogeneidad en el análisis, se aplicaron técnicas de mejora de imágenes, como el ajuste de brillo y contraste, eliminación de ruido, y redimensionamiento a un tamaño estándar. Adicionalmente, se utilizaron métodos de aumento de datos, como rotaciones, reflejos y escalados, para aumentar la cantidad de datos disponibles y mejorar la robustez del modelo.

En la fase de diseño e implementación del modelo, se desarrolló una red neuronal convolucional (CNN) personalizada, basada en arquitecturas previamente probadas en el análisis de imágenes médicas. Además, se exploró el uso de aprendizaje por transferencia, adaptando modelos preentrenados como ResNet o EfficientNet para la clasificación de las imágenes indeterminadas. El modelo fue entrenado utilizando los conjuntos de datos preprocesados, optimizando hiperparámetros clave como la tasa de aprendizaje, el número de épocas y el tamaño del lote para maximizar su desempeño.

Con el fin de garantizar la robustez del modelo, realizamos una evaluación rigurosa mediante una división estratificada de los datos y el cálculo de métricas de desempeño estándar como precisión, sensibilidad, especificidad y AUC. Además, llevamos a cabo una comparación con modelos de referencia para evaluar las mejoras obtenidas.

Finalmente, en la etapa de análisis de resultados, se interpretaron los hallazgos obtenidos, destacando el desempeño del modelo en la clasificación de imágenes indeterminadas. También se discutieron las limitaciones del enfoque y las posibles aplicaciones del modelo en entornos clínicos, proponiendo estrategias para su implementación práctica y su integración en sistemas de diagnóstico asistido por inteligencia artificial.

Esta metodología integral permitió abordar de manera efectiva el desafío de clasificar imágenes indeterminadas de cáncer cutáneo, contribuyendo al avance de las herramientas tecnológicas en el diagnóstico médico.

Fundamentación Teórica

El presente trabajo, se fundamenta en varios pilares teóricos que abarcan la medicina, el procesamiento de imágenes, el aprendizaje automático y el análisis de datos. Este marco proporciona una base sólida para la implementación de métodos avanzados dirigidos a abordar los retos que plantea la clasificación de imágenes indeterminadas en el contexto del cáncer de piel.

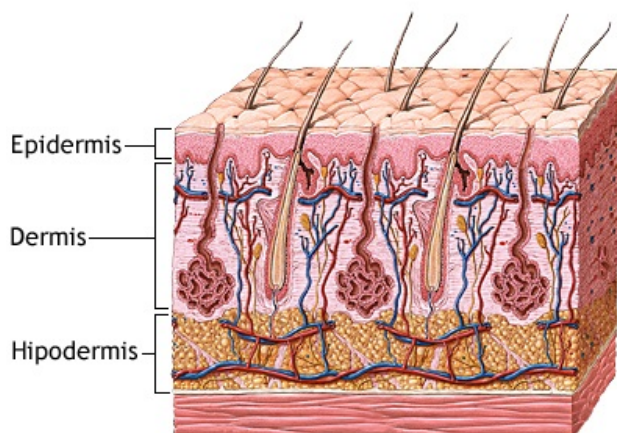
La Piel y su Importancia

Como barrera protectora, regulador térmico y órgano sensorial, la piel es indispensable para nuestra salud. Formada por tres capas principales epidermis, dermis e hipodermis, tal como lo muestra la Figura 2 este órgano, el más grande del cuerpo, cumple funciones vitales para nuestro bienestar. Cada una de estas capas tiene una estructura específica y desempeña un papel fundamental en la protección y el mantenimiento de la homeostasis del organismo.

La epidermis, la capa más superficial, compuesta por queratinocitos, funciona como una barrera protectora, mientras que los melanocitos le confieren color y la protegen de los rayos UV. La dermis, rica en colágeno y elastina, proporciona soporte y flexibilidad a la piel, y alberga estructuras como vasos sanguíneos, nervios y glándulas que regulan la temperatura y la hidratación. El tejido subcutáneo, o hipodermis, es la capa más interna y está compuesto principalmente por grasa y tejido conectivo. Este nivel proporciona aislamiento térmico, protección mecánica frente a impactos y almacenamiento de energía en forma de lípidos. La integridad de estas capas y su interacción armónica son esenciales para mantener la salud de la piel y del cuerpo en general (Benedetti, 2024).

Figura 2

Descripción de las Capas de Piel



Nota. El gráfico ilustra las diferentes capas de la piel, con sus respectivas descripciones, tomado de (*Capas de la piel: MedlinePlus enciclopedia médica ilustración, s.f.*).

El estudio de la piel, conocido como dermatología, es una especialidad médica que abarca el diagnóstico, tratamiento y prevención de enfermedades cutáneas. Los avances tecnológicos y científicos han permitido una mejor comprensión de la estructura, función y patologías de la piel, abriendo paso a herramientas de diagnóstico más precisas y tratamientos más efectivos.

Entre las herramientas más destacadas se encuentra la dermatoscopia, una técnica no invasiva que permite visualizar estructuras microscópicas de la piel no visibles a simple vista. Este método ha revolucionado el diagnóstico de enfermedades como el cáncer de piel, al facilitar la identificación de patrones específicos asociados a lesiones malignas o benignas. Además, el desarrollo de tecnologías de inteligencia artificial y análisis de datos ha ampliado las posibilidades de estudio, permitiendo clasificar imágenes y detectar anomalías con alta precisión (Martínez-González, 2021).

El estudio de la piel también tiene implicaciones en el ámbito preventivo. La investigación

sobre los efectos de factores externos, como la radiación UV, el envejecimiento y las condiciones ambientales, ha contribuido a diseñar estrategias para proteger la piel y prevenir enfermedades. La promoción de la protección solar y la educación sobre el autoexamen cutáneo son ejemplos claros de cómo la ciencia aplicada a la dermatología puede mejorar la salud pública.

Cáncer de Piel

El cáncer de piel es una de las enfermedades oncológicas más prevalentes a nivel mundial. Se divide principalmente en dos grandes categorías: cáncer de piel no melanoma (como el carcinoma basocelular y el carcinoma escamocelular) y melanoma maligno. De estos, el melanoma es el menos frecuente pero el más letal, siendo responsable de aproximadamente el 80 % de las muertes por cáncer cutáneo (A. F. P. López y cols., 2024).

El diagnóstico temprano de esta enfermedad es crucial, ya que las tasas de supervivencia disminuyen drásticamente en estadios avanzados. Sin embargo, el diagnóstico presenta desafíos significativos, especialmente cuando las imágenes dermatoscópicas muestran características ambiguas o indeterminadas. Este trabajo se centra en este tipo de imágenes, que representan un vacío crítico en las herramientas actuales de diagnóstico asistido.

Tipos de Cáncer de Piel

Cáncer de Piel No Melanoma. El carcinoma basocelular (CBC) y el carcinoma escamocelular (CCE) son los dos tipos principales de cáncer de piel no melanoma. El CBC, el más frecuente, se caracteriza por un crecimiento lento y local, mientras que el CCE es más agresivo y puede invadir tejidos profundos y metastatizar. Ambos suelen aparecer en áreas expuestas al sol y se presentan con diferentes características clínicas: el CBC como lesiones perladas o ulceradas, y el CCE como áreas de piel engrosada, escamosa o ulcerada. (Forero, 2024).

Melanoma Maligno. A pesar de ser menos frecuente, el melanoma es el tipo de cáncer de piel más agresivo y con mayor potencial metastásico. Se origina en los melanocitos y puede manifestarse como una lesión nueva o como una transformación maligna de un lunar preexistente. Su capacidad para diseminarse a otros órganos lo convierte en una enfermedad con un pronóstico reservado, por lo que su detección temprana es crucial para mejorar las opciones terapéuticas (A. F. P. López y cols., 2024).

Factores de Riesgo

El principal factor de riesgo para el desarrollo de cáncer de piel es la exposición excesiva a los rayos ultravioleta (UV), tanto solares como artificiales. Otros factores que incrementan el riesgo incluyen antecedentes familiares, fototipo cutáneo claro, presencia de numerosos nevos y un sistema inmunológico debilitado..

Además, ciertas condiciones ambientales y laborales, como la exposición prolongada a sustancias químicas carcinogénicas, también pueden aumentar el riesgo. En regiones tropicales o con altos niveles de radiación solar, como América Latina, la incidencia de cáncer de piel es particularmente elevada.

Prevención y Diagnóstico

La prevención del cáncer de piel se centra en reducir la exposición a la radiación UV mediante el uso de protector solar, ropa adecuada, y evitando la exposición directa al sol durante las horas de mayor intensidad. El autoexamen regular de la piel es una herramienta importante para detectar lesiones sospechosas en etapas tempranas (Brizuela Cruz, s.f.).

El diagnóstico se realiza a través de una combinación de examen físico, dermatoscopia y, en casos necesarios, biopsia. La dermatoscopia permite al dermatólogo observar características

internas de las lesiones, como la simetría, los bordes, los colores y los patrones específicos. En casos donde el diagnóstico no es claro, la biopsia proporciona una confirmación definitiva al analizar muestras de tejido bajo un microscopio.

Tratamiento

El tratamiento del cáncer de piel es altamente variable y depende del tipo histológico, estadio tumoral y características clínicas de la lesión. Los carcinomas basocelulares y escamosos suelen tratarse con cirugía excisional, mientras que el melanoma requiere un enfoque terapéutico más agresivo, que puede incluir cirugía, linfadenectomía, inmunoterapia y terapia dirigida. La dermatoscopia asistida por inteligencia artificial ha revolucionado el diagnóstico temprano del melanoma, permitiendo la identificación de lesiones sospechosas con mayor precisión y reduciendo la tasa de falsos negativos. La prevención del cáncer de piel sigue siendo fundamental y se basa en la protección solar adecuada, la realización de autoexámenes y la visita periódica al dermatólogo. (Martín Tinaquero, 2023).

Procesamiento de Imágenes Médicas

El procesamiento de imágenes médicas ha experimentado una revolución gracias a la aplicación de la inteligencia artificial. En el campo de la dermatología, las redes neuronales convolucionales han demostrado ser herramientas poderosas para el análisis de imágenes dermatoscópicas. Estas redes neuronales son capaces de aprender a reconocer patrones complejos en las imágenes, como la distribución del pigmento, la forma de los bordes y la presencia de estructuras vasculares, lo que permite una clasificación más precisa de las lesiones cutáneas y facilita la detección temprana del melanoma. Además, la inteligencia artificial puede ser utilizada para cuantificar características objetivas de las lesiones, lo que puede ayudar a predecir su evolución

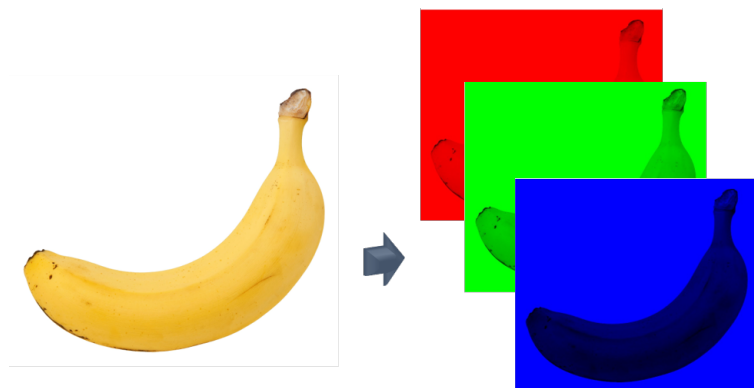
y a personalizar los tratamientos.(Martínez-González, 2021).

RGB: Modelo de Color

El modelo de color RGB es ampliamente utilizado en computación debido a su capacidad para representar una vasta gama de colores mediante la combinación de tres componentes básicos: rojo, verde y azul. Cada uno de estos componentes tiene un rango de intensidad que generalmente varía de 0 a 255, lo que permite definir hasta 16.7 millones de colores distintos (2563 combinaciones). Esta flexibilidad hace del modelo RGB una herramienta esencial en aplicaciones donde la representación visual precisa y detallada es crucial, como en la edición de imágenes, simulaciones gráficas y visualización de datos científicos. (Valentina, 2023).

Figura 3

Representación de una Imagen en el Espacio RGB



Nota. La figura muestra cómo una imagen se representa utilizando el modelo de color RGB, donde cada canal (rojo, verde y azul) contribuye a la composición final de la imagen. Autor, 2024.

Además de su uso en pantallas y dispositivos electrónicos, el modelo RGB es fundamental en el procesamiento de imágenes digitales (Figura 3). Al descomponer una imagen en sus canales de color individuales, es posible realizar análisis más específicos, como la detección de bordes, el reconocimiento de patrones o la segmentación de objetos en una escena. Esto lo convierte en una

base sólida para algoritmos avanzados en visión por computadora, aprendizaje profundo y sistemas de reconocimiento. Su versatilidad y capacidad para manejar datos visuales complejos explican por qué sigue siendo uno de los modelos de color más importantes en el ámbito tecnológico y científico. (Sanabria Codesal, 2023).

Aprendizaje Automático (Machine Learning)

El aprendizaje automático, o machine learning, es una rama de la inteligencia artificial que permite a las computadoras aprender de la experiencia sin ser programadas explícitamente. Al analizar grandes volúmenes de datos, estos sistemas pueden identificar patrones complejos y construir modelos predictivos que generalizan a nuevos datos. Esto ha permitido el desarrollo de aplicaciones innovadoras en diversos campos, como la visión por computadora, el procesamiento del lenguaje natural y la medicina. Por ejemplo, las redes neuronales convolucionales, un tipo de algoritmo de aprendizaje profundo, se utilizan para analizar imágenes médicas y detectar enfermedades como el cáncer con una precisión cada vez mayor. (Pertuz, 2022).

En esencia, el aprendizaje automático transforma datos en conocimiento útil, proporcionando a los sistemas la capacidad de adaptarse a nuevos escenarios y resolver problemas complejos. Su aplicabilidad abarca desde la clasificación de imágenes y el procesamiento del lenguaje natural hasta la predicción de eventos futuros en áreas como la medicina, la finanza, y la industria (Vega, Mora, Badilla, y cols., 2020).

Aprendizaje Supervisado

El aprendizaje supervisado es una técnica fundamental en el campo del machine learning que consiste en entrenar un modelo a partir de un conjunto de datos donde cada ejemplo está etiquetado con la respuesta correcta. Por ejemplo, para entrenar un modelo de clasificación de

imágenes, se le proporcionaría un conjunto de imágenes etiquetadas como “gato”ó “perro”. El modelo aprendería a identificar las características visuales que distinguen a un gato de un perro y, posteriormente, sería capaz de clasificar nuevas imágenes con alta precisión. Esta técnica se utiliza en una amplia variedad de aplicaciones, desde el reconocimiento de voz hasta la detección de fraudes.(Pertuz, 2022).

En el aprendizaje supervisado, el modelo procesa un conjunto de datos de entrenamiento compuesto por pares de entrada y salida conocidos. Durante este proceso, el algoritmo intenta minimizar la diferencia entre las predicciones generadas y las etiquetas reales mediante el ajuste de sus parámetros internos. Este ajuste se realiza utilizando funciones de pérdida y métodos de optimización, como la retropropagación en redes neuronales o el descenso por gradiente.

Una vez que un modelo de aprendizaje supervisado ha sido entrenado, es crucial evaluarlo con un conjunto de datos de prueba independiente. Este conjunto de datos, que no ha sido utilizado durante el entrenamiento, permite medir la capacidad del modelo para generalizar sus conocimientos y hacer predicciones precisas sobre nuevos datos. Si el modelo obtiene un alto rendimiento en el conjunto de prueba, se considera que ha aprendido las características relevantes de los datos y puede ser utilizado para realizar predicciones en situaciones reales. Por el contrario, si el modelo presenta un bajo rendimiento, es posible que esté sobreajustado, es decir, que haya memorizado los datos de entrenamiento en lugar de aprender las relaciones subyacentes.

Aprendizaje No Supervisado

En el aprendizaje no supervisado, el modelo trabaja directamente con las características de los datos sin asociarlas a salidas específicas. Los algoritmos procesan las entradas y buscan características comunes o patrones subyacentes. Esto puede implicar la agrupación de datos

similares (clustering), la reducción de dimensiones para simplificar su análisis, o la detección de anomalías dentro del conjunto (Pertuz, 2022).

El modelo no “aprende” de manera supervisada en términos de optimizar una función de pérdida, sino que se basa en métricas como la similitud entre datos o la densidad de puntos en el espacio de características. Esto permite la generación de resultados útiles para el análisis exploratorio y la comprensión de datos complejos.

Aprendizaje Profundo (Deep Learning)

El aprendizaje profundo, o deep learning, es un campo de la inteligencia artificial que ha revolucionado la forma en que las máquinas procesan información y toman decisiones. Inspirado en la arquitectura del cerebro humano, este enfoque emplea redes neuronales artificiales con múltiples capas para aprender representaciones jerárquicas de la información. Estas redes tienen la capacidad de extraer características avanzadas a partir de grandes volúmenes de datos, permitiéndoles realizar tareas complejas como el reconocimiento de imágenes, el procesamiento de lenguaje natural y la generación automática de texto. Los avances impulsados por el aprendizaje profundo han sido notables en áreas como la visión por computadora, la medicina y la robótica.(Rouhiainen, 2018).

El aprendizaje profundo se distingue de otros métodos de aprendizaje automático por su capacidad para extraer información directamente de los datos en su estado original, sin depender de características diseñadas previamente. Las redes neuronales profundas cuentan con múltiples capas ocultas que analizan los datos de forma progresiva, generando representaciones más complejas en cada etapa del proceso.

Por ejemplo, en una tarea de clasificación de imágenes, las primeras capas de la red pueden

aprender a detectar bordes y colores, mientras que las capas intermedias identifican formas o texturas, y las últimas capas comprenden estructuras completas, como objetos o rostros. Esta capacidad de “aprendizaje jerárquico” es una de las principales razones de su éxito en tareas complejas.

Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA) son sistemas computacionales inspirados en el funcionamiento del cerebro humano que han transformado la inteligencia artificial. Estas redes consisten en varias capas de nodos interconectados, donde cada nodo realiza operaciones matemáticas básicas. Durante el proceso de entrenamiento, las redes adquieren la habilidad de reconocer patrones complejos en los datos, lo que les permite desempeñar tareas como la identificación de imágenes, la traducción automática y la creación de música. Gracias a su capacidad para aprender de grandes volúmenes de datos, las redes neuronales han impulsado avances significativos en áreas como la medicina, las finanzas y la robótica. (Izaurieta y Saavedra, 2000; R. F. López y Fernández, 2008).

Estructura de una Red Neuronal Artificial

Una red neuronal artificial está formada por capas de nodos, comúnmente conocidos como “neuronas”. Estas capas se clasifican en tres categorías principales:

Capa de Entrada. Recibe los datos iniciales que se procesarán, representados en forma de vectores de características.

Capas Ocultas. Procesan la información a través de cálculos matemáticos complejos, utilizando pesos y funciones de activación para transformar los datos y extraer características relevantes.

Capa de Salida. Proporciona el resultado final del modelo, como una predicción, una clasificación o un valor continuo.

Cada conexión entre los nodos de una red neuronal tiene un valor numérico asociado que determina la intensidad de la señal que se transmite de un nodo a otro. Estos valores, denominados pesos, se ajustan de forma gradual durante el entrenamiento para que la red pueda aprender a reconocer patrones complejos en los datos (Acevedo, Serna, y Serna, 2017).

Funcionamiento y Entrenamiento

El entrenamiento de una red neuronal se basa en un proceso iterativo que incluye:

Propagación Hacia Adelante. Los datos de entrada pasan a través de las capas de la red, generando una predicción en la capa de salida.

Cálculo de Error. Se mide la discrepancia entre la predicción y el valor real utilizando una función de pérdida.

Retropropagación. El error se propaga hacia atrás a través de la red, ajustando los pesos de las conexiones para reducir el error en futuras iteraciones. Este proceso utiliza técnicas de optimización como el descenso por gradiente.

El aprendizaje puede ser supervisado, no supervisado o por refuerzo, dependiendo de la naturaleza de los datos y la tarea a resolver.

Tipos de Redes Neuronales

Existen múltiples tipos de redes neuronales, cada una diseñada para abordar problemas específicos:

Redes Neuronales Multicapa (MLP). Adecuadas para tareas generales de clasificación y regresión.

Redes Neuronales Convolucionales (CNN). Especializadas en el procesamiento de imágenes y datos espaciales, ideales para reconocer patrones visuales.

Redes Neuronales Recurrentes (RNN). Utilizadas para procesar datos secuenciales, como texto, audio o series temporales.

Redes Generativas (GAN). Capaces de crear nuevos datos basados en distribuciones aprendidas, aplicadas en generación de imágenes y creación de contenido.

Desarrollo del Modelo Basado en Deep Learning

El desarrollo del modelo de clasificación de imágenes indeterminadas de cáncer de piel basado en deep learning se centró en el uso de arquitecturas avanzadas de redes neuronales convolucionales (CNN) para abordar el desafío de diferenciar entre lesiones benignas y malignas en casos de diagnóstico complejo. Para este propósito, se implementaron las redes MobileNet, ResNet 101V2 y Xception, reconocidas por su eficiencia y rendimiento en tareas de visión por computadora tal como lo demuestra el estudio realizado por (Martínez-González, 2021).

Dataset

Para el desarrollo del modelo de clasificación de imágenes indeterminadas de cáncer de piel, se utilizó la base de datos pública ISIC (International Skin Imaging Collaboration), una de las colecciones más completas y reconocidas de imágenes dermatoscópicas. Esta base de datos es ampliamente utilizada en la investigación médica y en el desarrollo de modelos de aprendizaje profundo debido a la calidad y diversidad de sus imágenes, así como a la disponibilidad de anotaciones detalladas proporcionadas por expertos dermatológicos.

La base de datos ISIC se obtuvo directamente a través de su portal oficial, disponible en ISIC Archive (*ISIC — International Skin Imaging Collaboration*, s.f.). Esta plataforma proporciona acceso gratuito a miles de imágenes dermatoscópicas de alta resolución, etiquetadas con información relevante como el diagnóstico, características clínicas y metadata asociada.

Durante el proceso de obtención, se realizó una selección de imágenes relevantes, priorizando aquellas etiquetadas como indeterminadas o que representaban desafíos diagnósticos. Además, se incluyeron imágenes correspondientes a diversas clases, como lesiones benignas (por

ejemplo, nevos melanocíticos) y malignas (como melanomas), asegurando un balance adecuado para el entrenamiento del modelo.

Comprensión de los Datos

La comprensión de los datos es un paso esencial para garantizar que el modelo de aprendizaje profundo pueda extraer patrones significativos de las imágenes. Los datos de la base ISIC presentan características clave:

Variabilidad de Lesiones. La base de datos incluye múltiples tipos de lesiones cutáneas, como nevos melanocíticos, queratosis actínica, carcinomas y melanomas malignos. Esto permite entrenar modelos con una amplia representación de casos clínicos, lo que mejora su capacidad de generalización.

Metadatos Clínicos. Además de las imágenes, ISIC proporciona metadatos relevantes como la edad, género y localización anatómica del paciente. Aunque en este modelo no se utilizó directamente, esta información puede ser aprovechada en futuros trabajos para enriquecer el análisis mediante datos multimodales.

Anotaciones Expertas. Cada imagen está acompañada de anotaciones realizadas por dermatólogos expertos, lo que asegura la calidad y fiabilidad de las etiquetas diagnósticas. Estas etiquetas son esenciales para entrenar modelos supervisados y evaluar su desempeño.

Pre-Procesamiento de los Datos

La preparación de los datos, también conocida como preprocesamiento, es una etapa crucial en el desarrollo de cualquier modelo de aprendizaje automático o profundo. Este proceso garantiza que los datos utilizados sean consistentes, limpios y adecuados para el modelo, lo que mejora significativamente su capacidad de aprendizaje y precisión. En el caso de las imágenes de cáncer de

piel, este proceso permitió que el modelo se enfocara en las características relevantes de las lesiones, reduciendo la influencia de ruido o información irrelevante.

Etapas del Pre-Procesamiento

El preprocesamiento de las imágenes en este trabajo incluyó varias etapas clave que se describen a continuación:

Normalización. Los valores de los píxeles de las imágenes fueron escalados a un rango de 0 a 1. Este paso es fundamental en los modelos de deep learning, ya que normalizar los datos asegura que los valores de entrada estén en un rango uniforme, lo que facilita la convergencia durante el entrenamiento ver Figura 4.

Redimensionamiento. Todas las imágenes se ajustaron a dimensiones estándar (224x224 píxeles), necesarias para cumplir con los requisitos de entrada de las arquitecturas seleccionadas (MobileNet, ResNet 101V2, EfficientNet). Este paso garantizó la consistencia en las entradas al modelo y permitió aprovechar al máximo las capacidades de las redes preentrenadas.

Figura 4

Normalización de la Imagen



Imagen Original

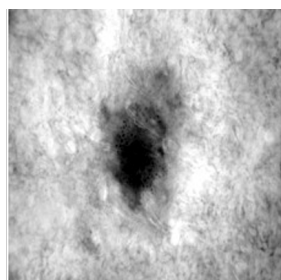


Imagen Normalizada

Nota. La figura ilustra el proceso de normalización de una imagen, que consiste en ajustar los valores de los píxeles para mejorar su rango dinámico y facilitar el análisis. Autor, 2024.

Aumento de Datos (Data Augmentation). Para enriquecer el conjunto de datos y prevenir el sobreajuste, se generaron nuevas versiones de las imágenes mediante transformaciones como:

- Zoom y escalado;
- Ajustes en el brillo y contraste;
- Traslaciones horizontales y verticales;

Estas transformaciones aumentaron la diversidad del conjunto de datos, lo que permitió que el modelo aprendiera a reconocer patrones de las lesiones desde diferentes perspectivas.

Desarrollo del Modelo

El desarrollo del modelo de clasificación de imágenes indeterminadas de cáncer cutáneo implicó un enfoque basado en el uso de arquitecturas avanzadas de redes neuronales convolucionales (CNN) preentrenadas: ResNet101, MobileNet y EfficientNet. Estas arquitecturas fueron seleccionadas por su capacidad para manejar problemas complejos de visión por computadora, como la clasificación de lesiones dermatoscópicas, y por su versatilidad al trabajar con datos médicos. A continuación, se describe el proceso detallado de desarrollo y entrenamiento del modelo.(Martínez-González, 2021).

Configuración de las Redes Pre-Entrenadas

Con el fin de aprovechar el conocimiento previo y acelerar el entrenamiento, se utilizaron redes neuronales preentrenadas en ImageNet como base. Estas redes, que han sido expuestas a una amplia variedad de imágenes naturales, proporcionan una representación inicial de las imágenes médicas. Posteriormente, se realizaron ajustes específicos a la arquitectura de la red, como modificar las capas finales y ajustar los hiperparámetros, para adaptarla a la tarea de clasificación de lesiones cutáneas. Esta estrategia de transferencia de aprendizaje nos permitió obtener un modelo

con un mejor rendimiento.

Parámetros de Entrenamiento para las Redes

El primer paso en la configuración fue cargar la red preentrenada utilizando bibliotecas de aprendizaje profundo como TensorFlow/Keras. Estas redes ya han aprendido características visuales generales (bordes, texturas, formas) a partir de grandes volúmenes de datos (como ImageNet), lo que las hace altamente efectivas para reutilizarlas en tareas específicas, en este caso, la clasificación de imágenes indeterminadas de cáncer de piel.

Parámetros Clave de Configuración. La configuración de parámetros de los modelos presentan las siguientes características:

`input_shape=(224, 224, 3)`: Define las dimensiones de las imágenes de entrada (224x224 píxeles y 3 canales RGB), asegurando compatibilidad con la arquitectura de la red.

`include_top=False`: Elimina las capas densas superiores diseñadas para clasificar las 1000 categorías de ImageNet. Esto permite personalizar las capas finales del modelo según las clases específicas de la tarea (lesión benigna o maligna).

`weights='imagenet'`: Indica que los pesos aprendidos en ImageNet deben ser reutilizados.

Esto proporciona un punto de partida sólido, reduciendo la necesidad de entrenar desde cero.

```

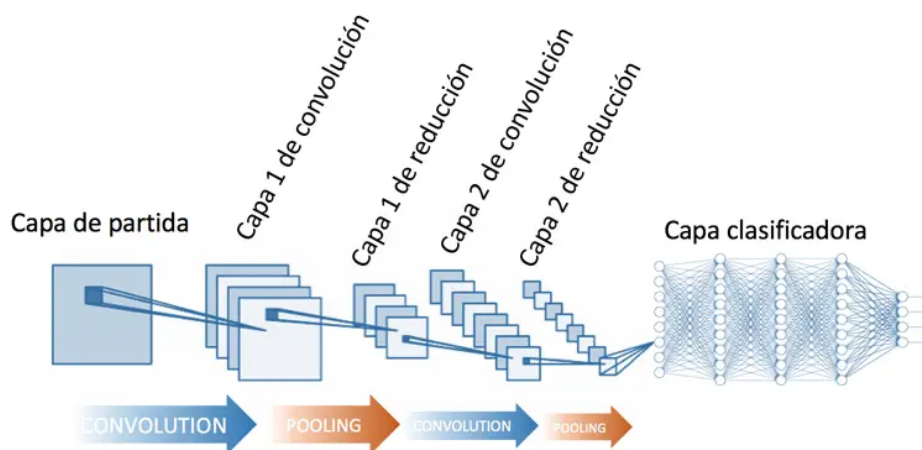
1 mdl = tf.keras.applications.Xception(
2     include_top=False,
3     weights='imagenet',
4     input_shape=(224, 224, 3),
5     pooling='avg',
6 )

```

Para observar una estructura similar, puedes referirte a la Figura 5 que ilustra una red neuronal convolucional, la cual incluye capas de entrada, convolución, agrupación y salida.

Figura 5

Red Convolucional



Nota. La imagen muestra la arquitectura de una red neuronal convolucional (CNN), destacando las capas principales: entrada, convolucionales, de reducción (pooling) y la capa final clasificadora.

Esta estructura es utilizada para procesar y clasificar imágenes mediante la extracción de características jerárquicas (Jauregui, 2023).

Ajuste de las Capas del Modelo. Una vez cargado el modelo preentrenado, se personalizó su estructura para adaptarlo a la tarea específica de clasificación. Este paso implicó extraer las capas convolucionales necesarias y definir las salidas intermedias:

Aislamiento de Capas Convolucionales. Se creó un modelo parcial (SplitModel) que incluyó solo las capas convolucionales hasta una capa específica, excluyendo cualquier capa redundante para la nueva tarea.

Habilitación del Entrenamiento. Las capas convolucionales de la red se configuraron como entrenables (`trainable=True`) para permitir el ajuste fino de los pesos según las

características del conjunto de datos de imágenes médicas.

```

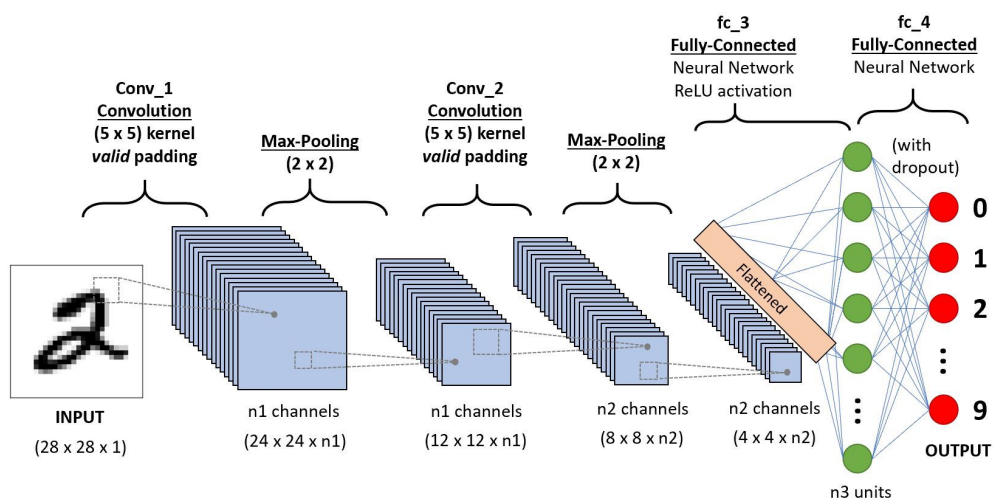
1 numlayers = len mdl.layers
2 SplitModel = Model(inputs=mdl.input,
3                   outputs=mdl.layers[numlayers-2].output,
4                   name='Xception')
5 SplitModel.trainable = True

```

Para observar una estructura similar a la red se puede observar en la Figura 6 se crea un modelo llamado SplitModel a partir de un modelo preexistente mdl, utilizando todas las capas excepto la última.

Figura 6

Ilustración de la Configuración de las Capas de la Red



Nota. Esta imagen muestra cómo se estructuran las capas en la red Xception, este enfoque es común cuando se desea utilizar una red neuronal preentrenada, para extraer características de las imágenes sin incluir la capa de clasificación final. Fuente. (*Xception Model and Depthwise Separable Convolutions*, 2019)

Adición de Capas Personalizadas. Para completar la red y adaptarla a la tarea de clasificación binaria, se añadieron capas densas personalizadas al modelo:

Global Average Pooling. Una capa de pooling promedio global se utilizó para comprimir la salida de las capas convolucionales a un vector unidimensional, lo que facilitó el entrenamiento y redujo la cantidad de parámetros.

Capa Densa de Salida. La salida final del modelo se definió con una capa Dense que contiene 2 neuronas (correspondientes a las clases benigna y maligna) y una función de activación softmax, que calcula las probabilidades de clasificación.

```

1 input_img = Input(shape=(224, 224, 3), name='Input')
2 out_SplitModel = SplitModel(input_img)
3 out_SplitModel = layers.GlobalAveragePooling2D()(out_SplitModel)
4 output = Dense(2, activation='softmax', name='Output')(out_SplitModel)
5 model = Model(input_img, output)

```

Compilación del Modelo. La compilación del modelo configuró su funcionamiento interno para optimizar el entrenamiento, en la Figura 7 se observa la estructura simulada para la ejecución.

Optimizador Adam. Este optimizador ajusta dinámicamente la tasa de aprendizaje durante el entrenamiento, lo que resulta ideal para problemas complejos como la clasificación de imágenes médicas.

Función de Pérdida Categorical_Crossentropy. Aunque la tarea es binaria, esta función se utiliza para calcular la pérdida en configuraciones multiclase, permitiendo mayor flexibilidad.

Métricas de Evaluación. La métrica de precisión (accuracy) monitoreó el desempeño del modelo en cada iteración.

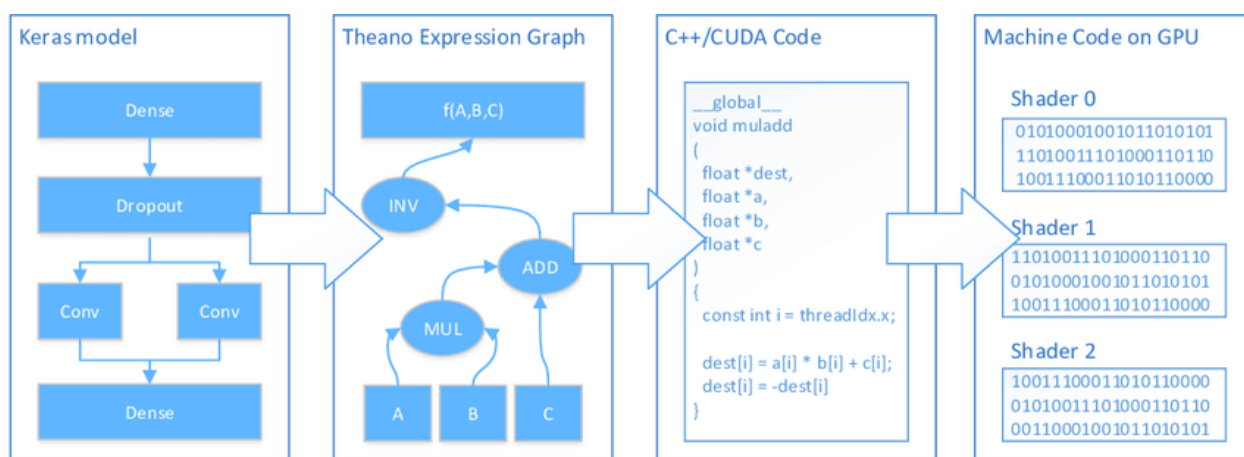
```

1 optimizer = Adam(1e-3) # 1e-5
2 model.compile(optimizer=optimizer,
3               loss='categorical_crossentropy',
4               metrics=['accuracy'])

```

Figura 7

Ilustración de la Compilación del Modelo



Nota. Esta imagen ilustra el flujo de trabajo de un modelo de aprendizaje profundo implementado en Keras. Se muestra cómo la estructura del modelo es representada como un grafo de expresiones en Theano, convertida a código en C++/CUDA, y finalmente traducida a código de máquina para su ejecución eficiente en GPU. Fuente. (Team, s.f.)

Entrenamiento del Modelo. Finalmente, el modelo fue entrenado utilizando el conjunto de datos segmentado en entrenamiento, validación y prueba. A lo largo de 50 épocas, el modelo ajustó sus pesos con el objetivo de reducir la pérdida y mejorar la precisión, empleando estrategias de aumento de datos y seguimiento constante de su rendimiento.

```

1 history = model.fit(train_batches,
2                     steps_per_epoch=steps_train,

```

```
3         validation_data=validation_batches,  
4         validation_steps=steps_validation,  
5         epochs=50)
```

El aprendizaje por transferencia es fundamental en este proyecto. Al iniciar el entrenamiento con una red preentrenada en una base de datos masiva, transferimos el conocimiento adquirido en tareas generales a la tarea específica de clasificación de lesiones cutáneas. Esta estrategia no solo acelera el proceso de aprendizaje, sino que también mejora la capacidad del modelo para extraer características relevantes de las imágenes médicas. La personalización de las últimas capas de la red y la incorporación de mecanismos de regularización garantizan un modelo óptimo para la tarea en cuestión.

Análisis de Resultados

En este capítulo se exponen los resultados obtenidos del entrenamiento y validación de diversas arquitecturas de redes neuronales profundas para la clasificación de imágenes indeterminadas de cáncer de piel. Se analiza el desempeño de estas redes en la tarea de identificar y clasificar las imágenes como benignas o malignas, junto con una descripción de los datos de entrenamiento y validación empleados en el estudio.

MobileNet

El análisis de los resultados del modelo MobileNet se basa en las métricas de evaluación, la curva de aprendizaje (precisión y pérdida) y la curva ROC. A continuación, se detalla el comportamiento del modelo:

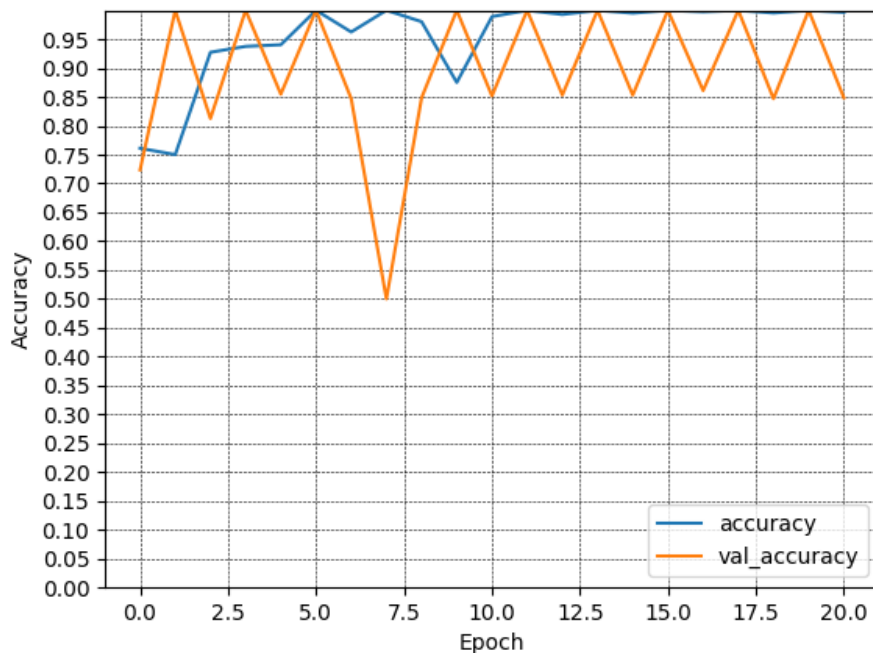
Precisión y Validación Durante el Entrenamiento

En la gráfica de precisión, tal como se muestra en la Figura 8 , se observa un buen desempeño inicial en el conjunto de entrenamiento, con valores que oscilan alrededor de 0.90 después de las primeras épocas. Sin embargo, la precisión en el conjunto de validación muestra una alta fluctuación, alcanzando valores cercanos a 0.95 en algunas épocas y cayendo drásticamente en otras, lo que indica un comportamiento inconsistente.

Este comportamiento sugiere que el modelo podría estar sobreajustándose a los datos de entrenamiento, ya que no logra generalizar bien a los datos de validación. Las oscilaciones en la validación pueden deberse a una falta de diversidad en los datos de entrenamiento o a la necesidad de técnicas adicionales de regularización, como dropout o ajuste del learning rate.

Figura 8

Gráfica de Precisión y Validación para el Modelo MobileNet



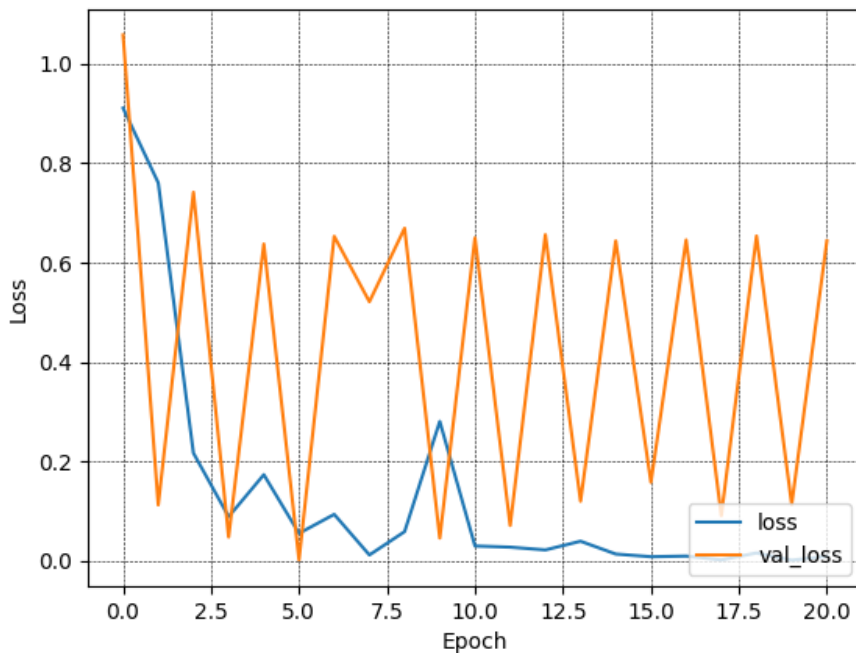
Nota. El gráfico muestra el desempeño del modelo en términos de precisión durante el entrenamiento y validación. Autor, 2024.

Perdida Durante el Entrenamiento

La gráfica presentada en la Figura 9, muestra la evolución de la pérdida durante el entrenamiento y la validación del modelo MobileNet a lo largo de 20 épocas. Se observa que la pérdida del conjunto de entrenamiento (línea azul) disminuye consistentemente, lo que indica que el modelo está aprendiendo de manera efectiva a ajustar sus parámetros para minimizar el error en los datos de entrenamiento. Sin embargo, la pérdida en el conjunto de validación (línea naranja) presenta un comportamiento fluctuante y con valores significativamente más altos que los de la pérdida de entrenamiento. Esto podría ser un indicativo de sobreajuste, donde el modelo se adapta demasiado a los datos de entrenamiento y pierde capacidad para generalizar en datos nuevos.

Figura 9

Pérdida Durante el Entrenamiento para el Modelo MobileNet



Nota. El gráfico muestra la evolución de la pérdida del modelo MobileNet durante el entrenamiento y validación. Autor, 2024.

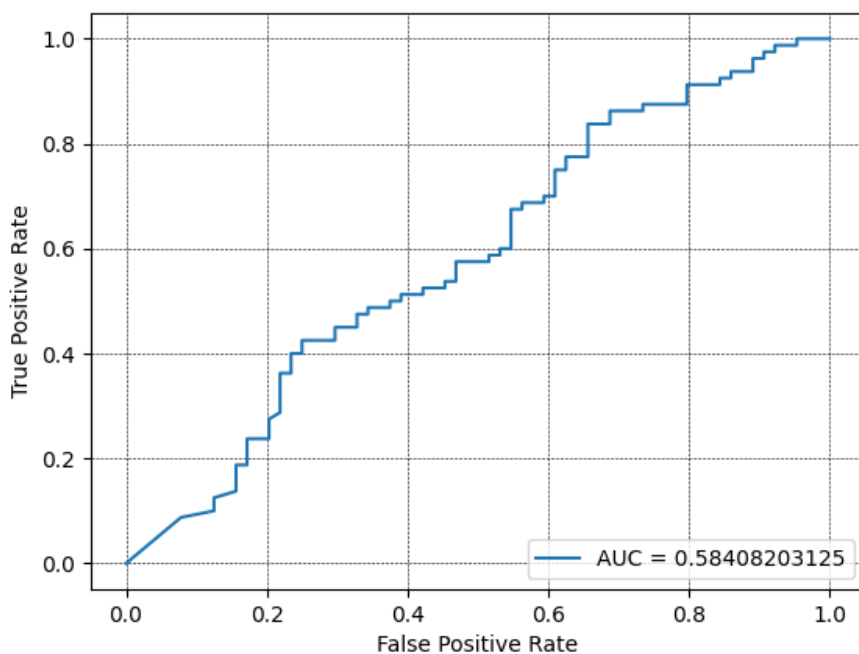
Además, las fluctuaciones en la pérdida de validación sugieren que el modelo enfrenta dificultades para estabilizarse durante la validación, posiblemente debido a la variabilidad en el conjunto de datos de validación o a una configuración de hiperparámetros que podría requerir ajustes adicionales. Este comportamiento resalta la necesidad de implementar técnicas para mejorar la capacidad de generalización del modelo, como el uso de regularización, reducción de la complejidad del modelo o la incorporación de más datos representativos para la validación. Estos ajustes podrían ayudar a reducir la brecha entre las curvas de pérdida de entrenamiento y validación, optimizando el rendimiento del modelo.

Curva ROC y Curva AUC

La curva ROC muestra un área bajo la curva (AUC) de 0,58, que es ligeramente superior al azar (0,50), pero lejos de ser ideal para un modelo destinado a la clasificación médica, tal como se observa en la Figura 10. Esto indica que el modelo tiene dificultades para distinguir eficazmente entre clases (benigna y maligna), probablemente debido a la falta de información clara o patrones consistentes en los datos.

Figura 10

Curva ROC (Receiver Operating Characteristic) para el Modelo MobileNet



Nota. El gráfico muestra la curva ROC del modelo MobileNet, representando su capacidad discriminativa entre clases. Autor, 2024.

Xception

El modelo Xception fue entrenado y evaluado utilizando las gráficas de pérdida, precisión y la curva ROC, lo que permitió analizar su desempeño en la tarea de clasificación de imágenes. A

continuación, se presentan los resultados clave y el análisis correspondiente.

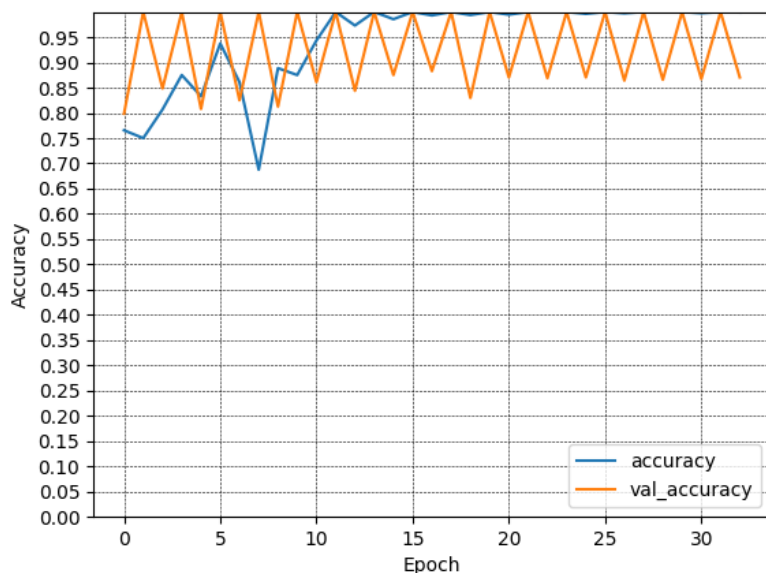
Precisión y Validación Durante el Entrenamiento

La gráfica muestra la precisión del modelo Xception durante el entrenamiento y la validación a lo largo de 30 épocas, tal como se observa en la Figura 11. Se observa que la precisión en el conjunto de entrenamiento (línea azul) mejora rápidamente en las primeras épocas, alcanzando valores cercanos al 95 %. Sin embargo, presenta pequeñas fluctuaciones a medida que avanza el entrenamiento, lo que podría indicar ligeros ajustes en los pesos del modelo para optimizar el rendimiento en los datos de entrenamiento. Por otro lado, la precisión del conjunto de validación (línea naranja) también alcanza valores altos desde las primeras épocas y mantiene un comportamiento más estable en comparación con la precisión de entrenamiento, lo que sugiere una adecuada generalización del modelo.

A pesar de los buenos resultados generales, la gráfica muestra una ligera discrepancia entre las precisiones de entrenamiento y validación en algunas épocas, lo que podría ser un indicativo de sobreajuste moderado. Sin embargo, la proximidad de ambas curvas hacia el final del entrenamiento indica que el modelo logra un balance razonable entre aprendizaje y generalización. Estos resultados reflejan que el modelo Xception es capaz de mantener un desempeño robusto, pero sería recomendable investigar técnicas adicionales, como la regularización o la reducción de la complejidad, para minimizar las fluctuaciones y mejorar aún más la estabilidad del modelo.

Figura 11

Gráfica de Precisión y Validación para el Modelo Xception



Nota. El gráfico muestra el desempeño del modelo Xception en términos de precisión durante el entrenamiento y validación. Autor, 2024.

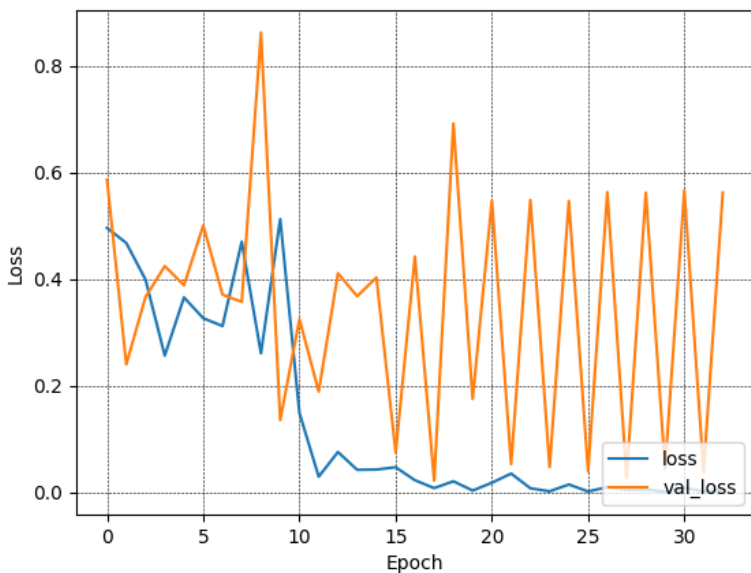
Perdida Durante el Entrenamiento

La gráfica de pérdida, como se observa en la Figura 12 muestra que la pérdida en el conjunto de entrenamiento disminuyó progresivamente a lo largo de las épocas, estabilizándose cerca de valores bajos hacia el final. Esto indica que el modelo fue capaz de aprender patrones consistentes en los datos de entrenamiento. Sin embargo, la pérdida en el conjunto de validación presentó fluctuaciones significativas, con picos altos que persisten incluso en las últimas épocas.

Estas oscilaciones en la pérdida de validación sugieren problemas de sobreajuste, donde el modelo aprende excesivamente los detalles y el ruido del conjunto de entrenamiento, dificultando su capacidad para generalizar a nuevos datos. Este problema podría ser mitigado utilizando técnicas de regularización como abandono o reducción de la complejidad del modelo.

Figura 12

Gráfica de Pérdida Durante el Entrenamiento para el Modelo Xception



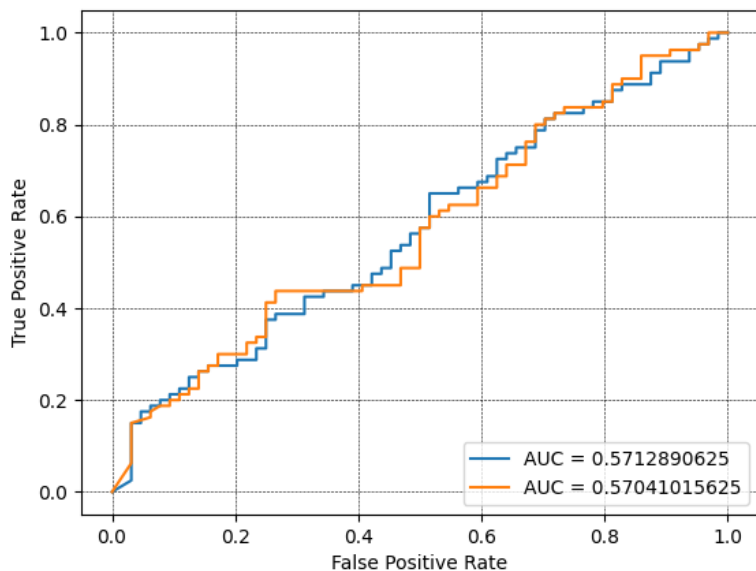
Nota. El gráfico muestra la evolución de la pérdida del modelo Xception durante el entrenamiento y validación. Autor, 2024.

Curva ROC y Curva AUC

La gráfica muestra la curva ROC (Receiver Operating Characteristic) del modelo Xception, mide su capacidad para diferenciar entre clases positivas (malignas) y negativas (benignas), tal como lo muestra la Figura 13. Los valores del área bajo la curva (AUC) son de 0.571 y 0.570 para dos configuraciones diferentes del modelo, indicando un desempeño similar en ambas. Estos valores de AUC sugieren que el modelo tiene una capacidad discriminativa moderadamente baja, apenas superior al azar. La proximidad de las curvas a la diagonal refleja que el modelo enfrenta dificultades para clasificar correctamente los datos con alta precisión.

Figura 13

Curva ROC (Receiver Operating Characteristic) para el Modelo Xception



Nota. El gráfico representa el comportamiento del modelo de los verdaderos positivos y los falsos positivos. Autor, 2024.

Este resultado destaca la necesidad de optimizar el modelo para mejorar su capacidad discriminativa. Algunas estrategias incluyen el ajuste de hiperparámetros, el uso de técnicas de regularización para evitar el sobreajuste y el incremento del tamaño o diversidad del conjunto de datos. Aunque el desempeño actual de Xception no alcanza niveles óptimos, los resultados obtenidos son una base útil para identificar posibles mejoras en la arquitectura y en los métodos de entrenamiento, con el objetivo de incrementar su aplicabilidad en tareas de clasificación médica como la detección de cáncer de piel.

ResNet101V2

El modelo ResNet101V2 fue entrenado y evaluado utilizando las gráficas de pérdida, precisión y la curva ROC. A continuación, se presenta el análisis detallado de los resultados

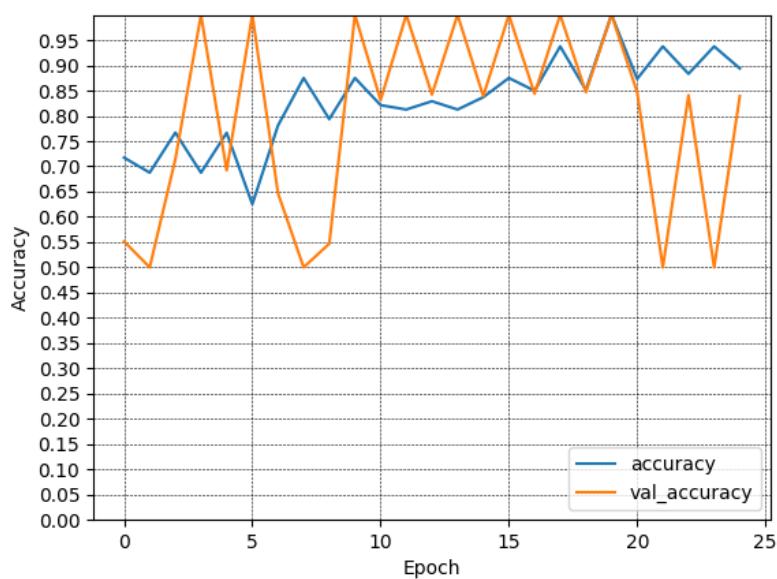
obtenidos:

Precisión y Validación Durante el Entrenamiento

En la gráfica de precisión, el modelo alcanza valores altos de precisión en el conjunto de entrenamiento, estabilizándose alrededor del 95 % . La precisión en el conjunto de validación también alcanza valores altos (cerca del 90 %), pero presenta oscilaciones importantes, especialmente en las primeras épocas y en algunas etapas posteriores del entrenamiento.

Figura 14

Gráfica de Precisión y Validación para el Modelo ResNet



Nota. El gráfico muestra el desempeño del modelo ResNet en términos de precisión durante el entrenamiento y validación. Autor, 2024.

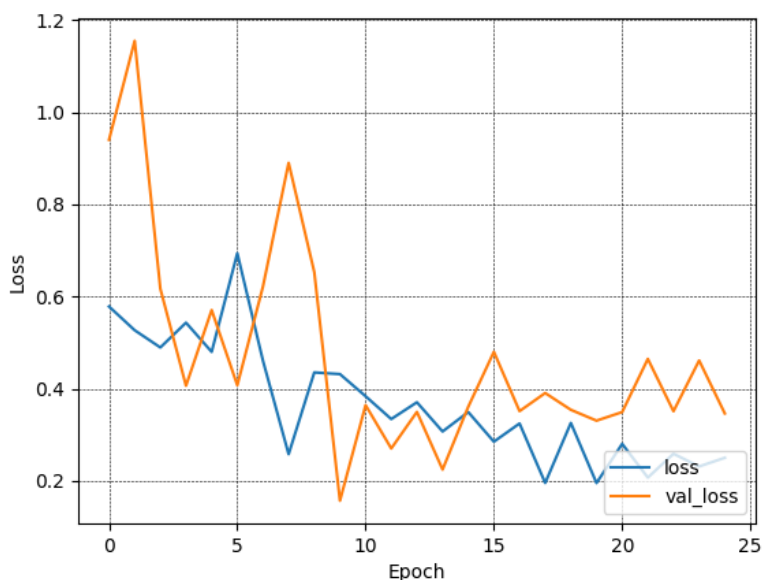
La diferencia entre la precisión del entrenamiento y la validación es un indicio de que el modelo podría estar sobreajustándose a los datos de entrenamiento. Esto podría corregirse utilizando técnicas de regularización como abandono o aumentando el conjunto de datos de entrenamiento para mejorar su capacidad de generalización (Figura 14).

Perdida Durante el Entrenamiento

El análisis de las curvas de pérdida revela un comportamiento típico en el entrenamiento de modelos de aprendizaje automático, tal como se observa en la Figura 15. La disminución continua de la pérdida en el conjunto de entrenamiento es esperada, ya que el modelo se ajusta progresivamente a los datos de entrenamiento. Sin embargo, las fluctuaciones observadas en la pérdida del conjunto de validación, especialmente en las primeras épocas, podrían indicar problemas de sobreajuste o variabilidad en los datos de validación.

Figura 15

Gráfica de Perdida para el Modelo ResNet



Nota. El gráfico muestra la evolución de la pérdida del modelo ResNet durante el entrenamiento y validación. Autor, 2024

El sobreajuste ocurre cuando el modelo se vuelve demasiado complejo y captura el ruido presente en los datos de entrenamiento en lugar de aprender las características generales. La variabilidad en el conjunto de validación puede dificultar la evaluación del desempeño del modelo,

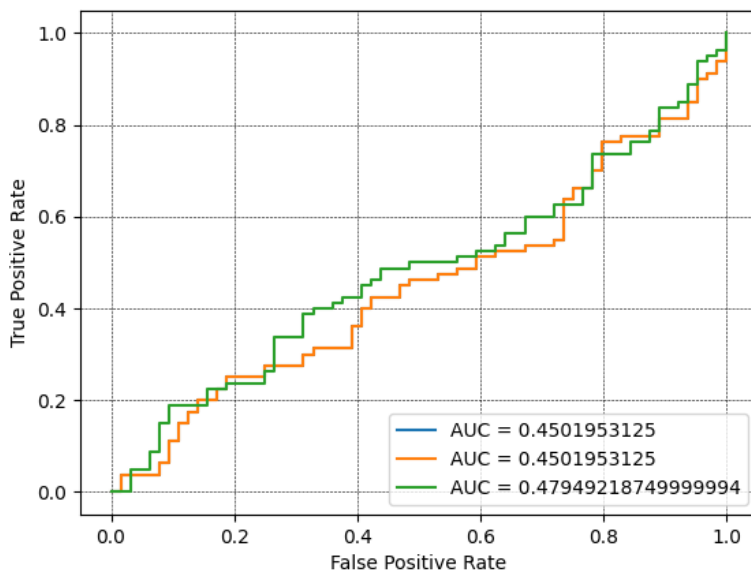
ya que pequeñas variaciones en los datos pueden generar grandes cambios en la pérdida. A pesar de estas posibles dificultades, la tendencia general de disminución en ambas curvas sugiere que el modelo está aprendiendo.

Curva ROC y Curva AUC

El análisis de la curva ROC revela un área bajo la curva (AUC) de 0,45, lo cual indica una capacidad discriminatoria extremadamente baja del modelo. Este valor se encuentra significativamente por debajo de lo esperado para un modelo de clasificación efectivo. Un AUC cercano a 0,5 sugiere que el modelo está realizando predicciones aleatorias y no está aportando ninguna información útil para la tarea de clasificación, tal como se observa en la Figura 16.

Figura 16

Curva ROC (Receiver Operating Characteristic) para el Modelo ResNet



Nota. El gráfico muestra la curva ROC del modelo ResNet, representando su capacidad discriminatoria entre clases. Autor, 2024

Varios factores podrían estar influyendo en este bajo desempeño. Un desequilibrio en la

distribución de clases puede llevar a que el modelo se sesgue hacia la clase mayoritaria, lo que dificulta la identificación de la clase minoritaria. Además, la calidad y representatividad de las características utilizadas para entrenar el modelo son cruciales. Si las características no son informativas o están altamente correlacionadas, el modelo tendrá dificultades para aprender los patrones subyacentes que permiten discriminar entre las clases.

Definiciones para el Calculo de las Gráficas

Se utilizaron los siguientes cálculos para las gráficas de barras de error y dispersión:

Barras de Error

El intervalo de confianza se calculó con la fórmula:

$$IC = z \cdot \frac{\sigma}{\sqrt{n}}$$

Puntos de Datos

Se representaron los puntos de las métricas para cada red con diferentes marcadores:

- MobileNet: círculo (○);
- Xception: diamante (D);
- ResNet101V2: cuadrado (s);

Líneas de Referencia

Se añadieron las siguientes líneas de referencia:

- Media (μ): Línea negra discontinua;
- Máximo (máx): Línea naranja punteada;
- Mínimo (mín): Línea púrpura punteada;

Desempeño de las Redes

Los resultados obtenidos a partir de las métricas de clasificación para los modelos MobileNet, Xception y ResNet101V2 reflejan el desempeño de cada red en la tarea de clasificación de imágenes de cáncer cutáneo, tal como se observa en la Tabla 2. A continuación, se presentan los puntos clave y el análisis correspondiente.

Tabla 2

Resultados de la Clasificación de los Modelos

Tipo Red	Exactitud	Sensibilidad	Especificidad	AUC
MobileNet	0.59	0.90	0.20	0.584
Xception	0.56	0.83	0.21	0.571
ResNet101V2	0.486	0.77	0.12	0.47

Nota. La tabla presenta los resultados de clasificación obtenidos para cada uno de los modelos evaluados. Se comparan métricas clave como la exactitud, sensibilidad, especificidad y área bajo la curva (AUC).

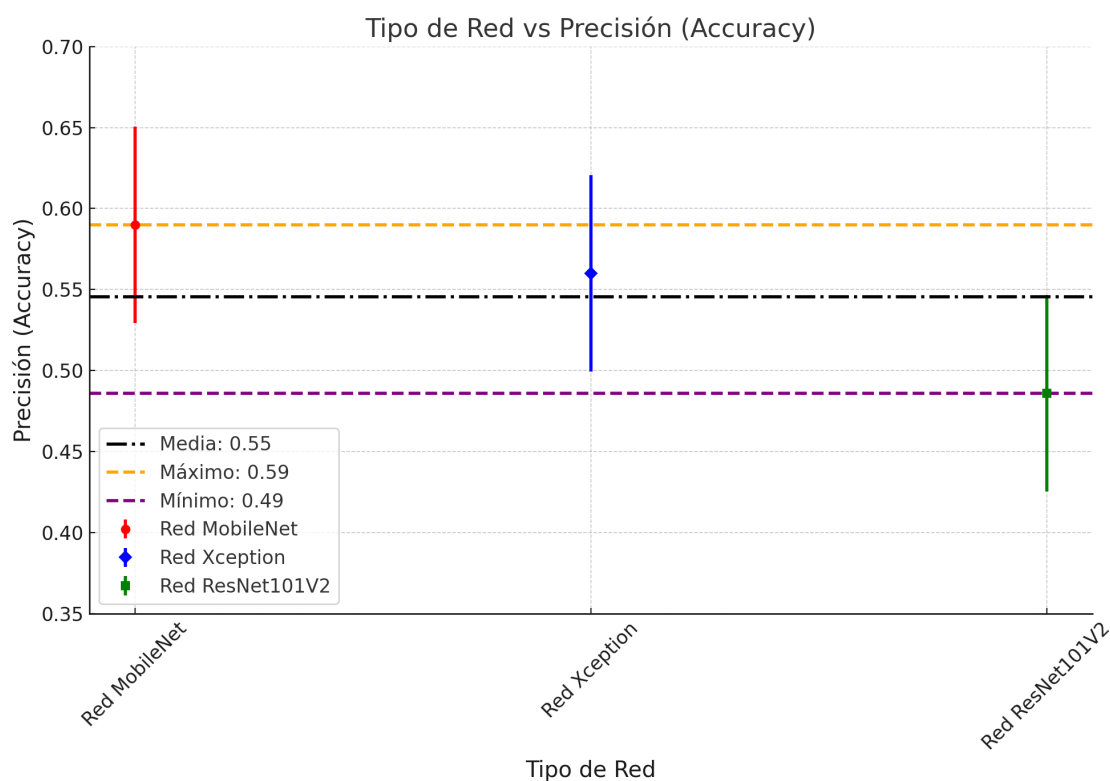
Análisis de la Exactitud (ACC)

Realizando el análisis de la exactitud (Acc), tal como lo muestra la Figura 17, para los tres modelos evaluados: MobileNet, Xception y ResNet101V2. El modelo MobileNet se posiciona como la arquitectura con el mejor desempeño, alcanzando un valor de precisión de 0.59, el más alto entre los tres modelos. Este resultado está claramente por encima del promedio general (0,55) y demuestra la capacidad de MobileNet para capturar patrones relevantes en los datos de clasificación. Xception, con una precisión de 0.56, también se encuentra por encima del promedio y presenta un

desempeño competitivo en comparación con MobileNet, lo que lo convierte en una alternativa viable en ciertas situaciones. Por otro lado, ResNet101V2 tiene el peor desempeño, con una precisión de 0.49 , lo que está por debajo del promedio y del mínimo esperado, indicando que este modelo enfrenta mayores dificultades para generalizar los patrones en el conjunto de datos.

Figura 17

Desempeño de la Exactitud de las Redes



Nota. El gráfico compara el desempeño en términos de exactitud de las redes MobileNet, Xception y ResNet. Autor, 2024.

Es importante destacar las diferencias en la dispersión entre los modelos. Mientras que MobileNet muestra un rango más estrecho y estable alrededor de su valor de precisión, Xception presenta una variación mayor, lo que puede reflejar problemas de consistencia en el rendimiento del modelo. ResNet101V2, además de tener el valor más bajo, muestra una dispersión considerable

hacia valores inferiores, lo que refuerza la percepción de su bajo desempeño. En general, la gráfica resalta la superioridad de MobileNet para la tarea específica de clasificación de imágenes de cáncer cutáneo, aunque su precisión máxima aún deja espacio para mejoras adicionales.

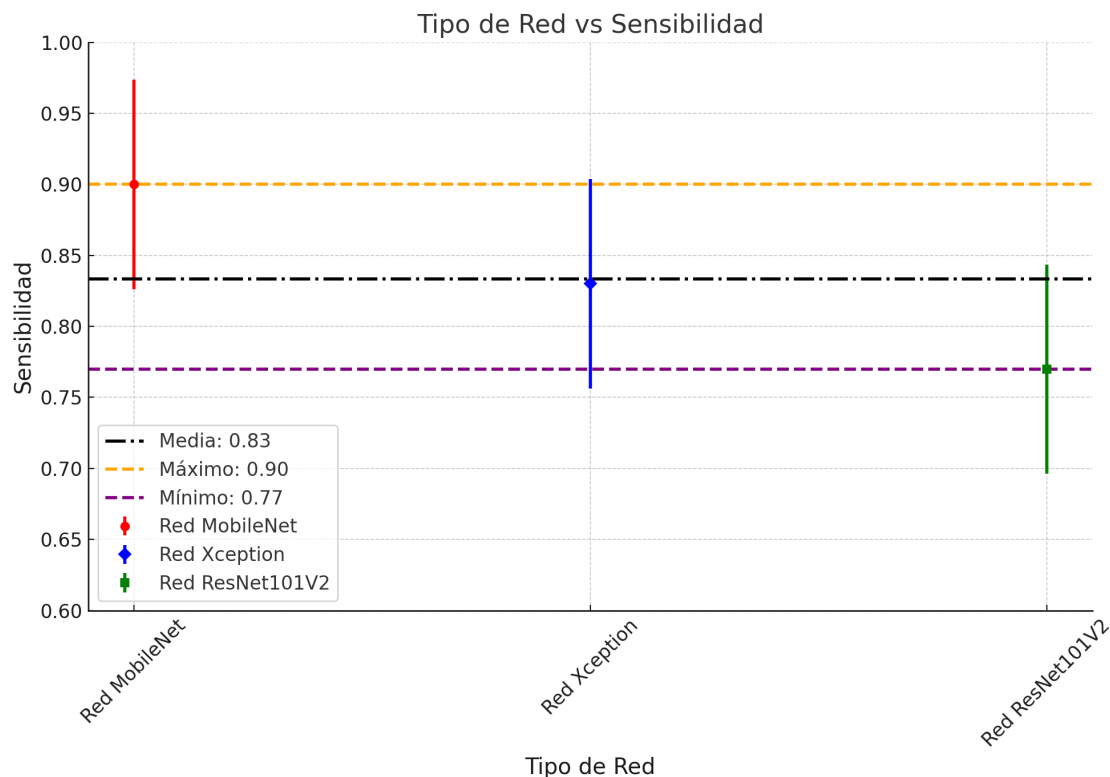
Análisis de la Sensibilidad (Sens)

La sensibilidad de los tres modelos (MobileNet, Xception y ResNet101V2), evidenciada por la Figura 18, mostró que la red MobileNet se destaca con el valor más alto de sensibilidad (0.90), reflejando su capacidad superior para detectar correctamente los casos positivos (lesiones malignas). Este resultado es crucial en el contexto médico, ya que priorizar la detección temprana de casos malignos es fundamental para el tratamiento exitoso. Xception, con una sensibilidad de 0.83 , también muestra un desempeño sólido, aunque ligeramente inferior al de MobileNet. Por otro lado, ResNet101V2 presenta la sensibilidad más baja (0.77), lo que indica que este modelo tuvo dificultades para identificar de manera consistente los casos positivos, reduciendo su utilidad en un entorno clínico donde la precisión en la detección de malignidades es esencial.

La media general de sensibilidad entre los tres modelos es de 0.83 , lo que refleja un desempeño aceptable, aunque ResNet101V2 queda por debajo del promedio. La dispersión observada en la gráfica indica una mayor variabilidad en los resultados de ResNet101V2, mientras que MobileNet tiene un rango más ajustado y consistente en torno a su valor de sensibilidad. Estos resultados confirman que MobileNet no solo es el modelo más efectivo para detectar lesiones malignas, sino también el más estable en términos de sensibilidad. Sin embargo, para un equilibrio completo en un entorno clínico, es necesario complementar la alta sensibilidad con una mejor especificidad, un área donde MobileNet y los otros modelos necesitan mejoras.

Figura 18

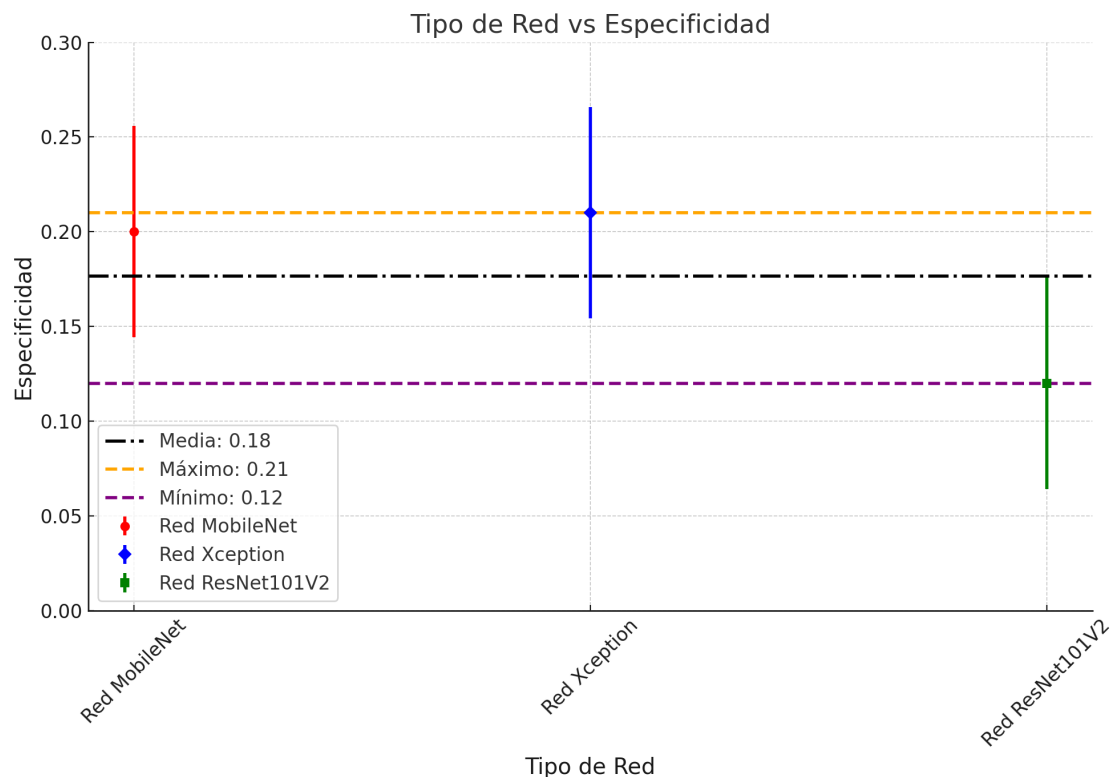
Desempeño de la Sensibilidad de las Redes



Nota. El gráfico compara el desempeño en términos de sensibilidad de las redes MobileNet, Xception y ResNet. Autor, 2024.

Análisis de la Especificidad (Spec)

Los tres modelos evaluados (MobileNet, Xception y ResNet101V2) tuvieron dificultades para distinguir correctamente las lesiones benignas en imágenes dermatoscópicas, tal como se evidencia en la Figura 19 con una especificidad promedio muy baja (0.18). Si bien Xception (0.21) y MobileNet (0.20) obtuvieron mejores resultados que ResNet101V2 (0.12), ninguno de ellos logró una tasa de falsos positivos aceptable, lo que sugiere la necesidad de explorar otras arquitecturas o técnicas para mejorar la precisión en este tipo de clasificación.

Figura 19*Desempeño de la Especificidad de los Modelos*

Nota. El gráfico compara el desempeño en términos de especificidad de las redes MobileNet, Xception y ResNet. Autor, 2024.

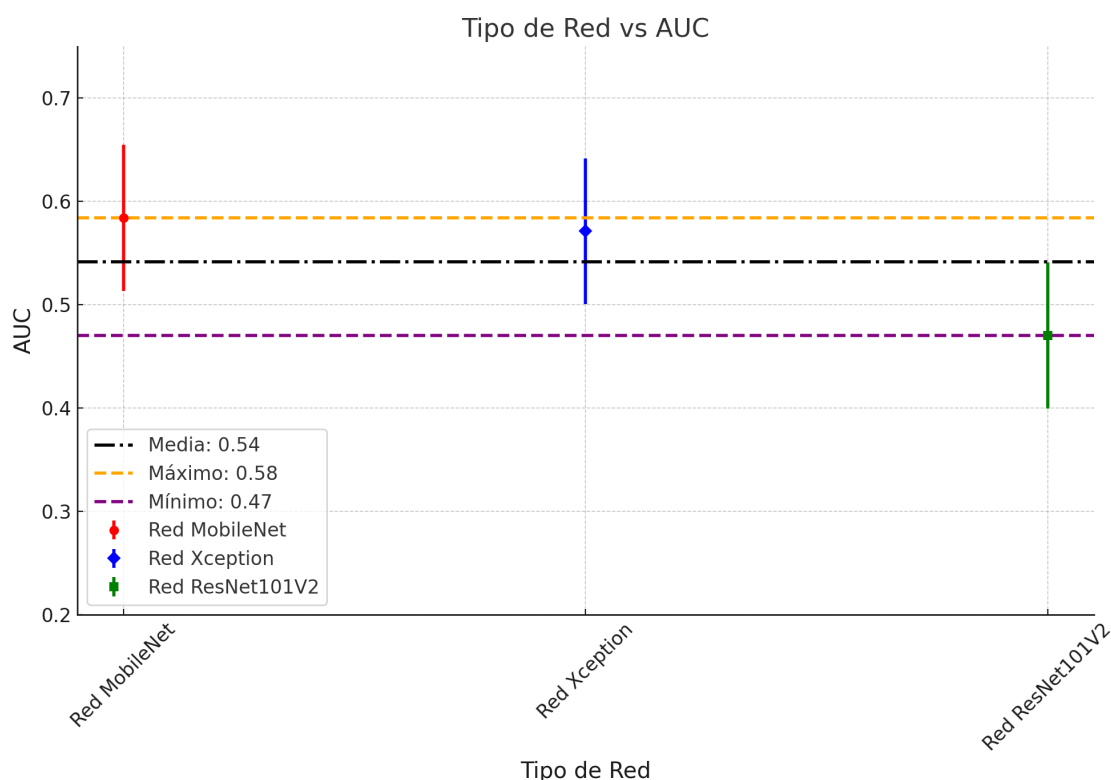
La baja especificidad observada en los tres modelos sugiere que tienden a clasificar incorrectamente muchas lesiones benignas como malignas, lo cual podría generar una alta tasa de falsos positivos. Este comportamiento es preocupante en un contexto médico, ya que puede llevar a intervenciones innecesarias y estrés en los pacientes. Además, la dispersión en la gráfica indica que MobileNet y Xception presentan una mayor estabilidad en comparación con ResNet101V2, que muestra una variabilidad significativa hacia valores más bajos. Para mejorar la especificidad, sería necesario implementar técnicas de ajuste de umbral, aumentar el equilibrio en el conjunto de datos y utilizar enfoques híbridos para combinar las fortalezas de los modelos.

Análisis del Área Bajo la Curva (AUC)

En la Figura 20, se muestra un comparativo del desempeño de tres modelos de redes neuronales en una tarea de clasificación, utilizando el área bajo la curva ROC (AUC) como métrica. El AUC es un indicador de qué tan bien un modelo puede distinguir entre clases positivas y negativas.

Figura 20

Desempeño del Área Bajo la Curva de las Redes



Nota. El gráfico compara el desempeño del área bajo la curva (AUC) de las redes MobileNet, Xception y ResNet. Autor, 2024

En general, se observa una variabilidad considerable en el desempeño de los modelos. Xception presenta el mejor desempeño con un valor de AUC de 0.6, lo que sugiere una buena

capacidad para discriminar entre las clases. Por otro lado, MobileNet y ResNet101V2 muestran resultados más bajos, con valores de AUC de 0.47 y 0.5 respectivamente. Esto indica que estos dos últimos modelos tienen dificultades para diferenciar correctamente entre las clases. Además, las líneas de error asociadas a cada modelo sugieren un nivel de incertidumbre en los resultados, lo que podría estar relacionado con la variabilidad en los datos de entrenamiento o con la complejidad de la tarea de clasificación.

Evaluación Estadística del Desempeño de los Modelos

El análisis estadístico realizado en este proyecto permitió evaluar de manera objetiva el desempeño de los tres modelos de redes neuronales (MobileNet, Xception y ResNet101V2) en la clasificación de imágenes indeterminadas dermatoscópicas de lesiones cutáneas. A través de un análisis estadístico riguroso, se calcularon métricas como precisión, sensibilidad, especificidad y AUC para comparar de manera objetiva los modelos.

Definiciones Estadísticas

Para evaluar de manera rigurosa el rendimiento de los modelos de redes neuronales, se calcularon diversas métricas estadísticas. A continuación, se presentan las definiciones matemáticas de las medidas estadísticas utilizadas:

Media. La media, comúnmente llamada promedio, es una herramienta estadística que sintetiza un conjunto de datos al determinar su valor central o representativo. Este indicador es fundamental porque ofrece una perspectiva general sobre el desempeño esperado de cada modelo, evitando centrarse en casos particulares o valores extremos. Esta medida de tendencia central se define mediante la fórmula:

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

Donde:

- x_i : Valor de la observación i ;
- n : Número total de observaciones;

Desviación Estándar. La desviación estándar es una herramienta estadística que refleja la variabilidad o dispersión de un conjunto de datos con respecto a su media. Esta métrica cuantifica cómo se distribuyen los datos alrededor del promedio y se calcula mediante la ecuación:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}$$

Intervalo de Confianza. El intervalo de confianza al 95 % estima el rango dentro del cual se espera que se encuentre la verdadera media:

$$IC = z \cdot \frac{\sigma}{\sqrt{n}}$$

Donde:

- z : Valor crítico para un nivel de confianza del 95 % (1.96);
- σ : Desviación estándar;
- n : Número total de observaciones;

En entornos del mundo real, los intervalos de confianza ayudan a comprender la incertidumbre asociada con una métrica y evaluar la robustez de un modelo.

Análisis Estadístico

Esta sección compara el desempeño de tres arquitecturas de redes neuronales (MobileNet, Xception y ResNet101V2) en la tarea de clasificación de imágenes dermatoscópicas. Se calculó un conjunto de métricas de rendimiento (precisión, sensibilidad, especificidad y AUC) y se presentaron

estadísticas descriptivas detalladas para facilitar la comparación entre los modelos. Los resultados obtenidos permiten identificar el modelo con el mejor desempeño general y evaluar su robustez, tal como se observan en la Tabla 3.

Tabla 3

Estadísticas Detalladas para Cada Métrica y Red

Métrica	Red	Media	Desv.	Mín.	Max.	IC (95 %)
Accuracy	MobileNet	0.590	0.000	0.590	0.590	0.000
	Xception	0.560	0.000	0.560	0.560	0.000
	ResNet101V2	0.486	0.000	0.486	0.486	0.000
Sensibilidad	MobileNet	0.900	0.000	0.900	0.900	0.000
	Xception	0.830	0.000	0.830	0.830	0.000
	ResNet101V2	0.770	0.000	0.770	0.770	0.000
Especificidad	MobileNet	0.200	0.000	0.200	0.200	0.000
	Xception	0.210	0.000	0.210	0.210	0.000
	ResNet101V2	0.120	0.000	0.120	0.120	0.000
AUC	MobileNet	0.584	0.000	0.584	0.584	0.000
	Xception	0.571	0.000	0.571	0.571	0.000
	ResNet101V2	0.470	0.000	0.470	0.470	0.000

Nota. La tabla presenta un resumen estadístico detallado de cada métrica evaluada para las redes MobileNet, Xception y ResNet101V2. Se incluyen valores de media, desviación estándar, mínimos, máximos y el intervalo de confianza al 95 %.

Variabilidad

La variabilidad en los resultados, representada por la desviación estándar y el rango en la tabla 3, es nula en todas las métricas y redes evaluadas, lo que indica una completa consistencia en los valores obtenidos durante las evaluaciones. Esta ausencia de variabilidad sugiere que los modelos MobileNet, Xception y ResNet101V2 tuvieron un comportamiento homogéneo en términos de precisión, sensibilidad, especificidad y AUC en el conjunto de datos utilizados. Sin embargo, esta estabilidad también podría ser una señal de limitaciones en el diseño experimental, como un conjunto de datos poco diversos o un método de validación estático que no captura escenarios más variados. Por lo tanto, aunque la falta de dispersión garantiza resultados reproducibles, es importante incorporar datos más heterogéneos en futuros trabajos para evaluar mejor la robustez y generalización de los modelos.

Precisión de las Estimaciones

La precisión de las estimaciones, evaluada a través de los intervalos de confianza al 95 % en la tabla 3, es absoluta, ya que todos los valores de los intervalos son 0.000. Esto indica que las métricas calculadas (precisión, sensibilidad, especificidad y AUC) fueron completamente consistentes y no se mostraron variaciones en las pruebas realizadas. Si bien esto puede interpretarse como un alto nivel de estabilidad en los modelos MobileNet, Xception y ResNet101V2, también puede reflejar limitaciones en la diversidad del conjunto de datos o en el diseño del experimento. En condiciones ideales, los intervalos de confianza más amplios permitirían capturar la incertidumbre inherente en las estimaciones, especialmente al evaluar modelos en contextos más variados.

Comparación Entre Redes

La comparación entre las redes MobileNet, Xception y ResNet101V2, basada en la tabla 3, revela diferencias significativas en su desempeño. MobileNet se destaca como la mejor red en términos de precisión (0.590) y sensibilidad (0.900), lo que indica su eficacia para detectar correctamente las lesiones malignas y clasificar con mayor precisión en general. Xception, con valores ligeramente inferiores de precisión (0.560) y sensibilidad (0.830), se posiciona como una alternativa competitiva, aunque menos eficiente que MobileNet. Por otro lado, ResNet101V2 muestra el peor desempeño en todas las métricas, con una precisión de 0.486 , sensibilidad de 0.770 y especificidad de 0.120 , evidenciando dificultades tanto para identificar correctamente las lesiones benignas como malignas. Además, en términos de especificidad, Xception lidera con un 0.210 , mientras que MobileNet alcanza un valor similar (0.200) y ResNet101V2 queda rezagado. La métrica AUC, que mide la capacidad de discriminación, confirma esta tendencia, con MobileNet al frente (0.584) y ResNet101V2 como el menos efectivo (0.470). Estos resultados subrayan la superioridad de MobileNet como la red más equilibrada, mientras que Xception destaca como una alternativa viable, y ResNet101V2 requiere ajustes significativos para mejorar su rendimiento.

Conclusiones y Recomendaciones

Conclusiones

El presente trabajo evaluó tres modelos de redes neuronales profundas preentrenadas, MobileNet, Xception y ResNet101V2, para la tarea de clasificación de imágenes indeterminadas de cáncer cutáneo. Los resultados obtenidos proporcionan valiosas perspectivas sobre el desempeño de estas arquitecturas en un contexto médico complejo.

MobileNet demostró ser el modelo más prometedor, destacándose por su alta sensibilidad (0.9), lo que implica que el modelo es efectivo en la detección de lesiones malignas, minimizando los falsos negativos. Esta característica es crucial en aplicaciones clínicas, donde la prioridad es evitar que un caso maligno pase desapercibido. Sin embargo, la baja especificidad del modelo (0.2) indica que clasifica incorrectamente un número considerable de lesiones benignas como malignas, lo que puede llevar a un aumento en procedimientos médicos innecesarios.

Xception mostró un desempeño similar al de MobileNet, con una sensibilidad alta (0.83) y una precisión general comparable. No obstante, su especificidad (0.21) sigue siendo baja, reflejando las mismas dificultades para clasificar correctamente las imágenes benignas. Este comportamiento sugiere que, aunque Xception puede ser una alternativa válida, sus limitaciones lo hacen menos efectivo que MobileNet en esta tarea.

ResNet101V2 presentó el desempeño más bajo, con una precisión general de solo 0.486, sensibilidad de 0.77 y especificidad de 0.12. Estos resultados indican que el modelo no logró captar patrones discriminativos claros entre las clases, reflejando dificultades significativas tanto en la detección de lesiones malignas como benignas. Esto posiciona a ResNet101V2 como la opción

menos adecuada entre las arquitecturas evaluadas para esta tarea específica.

En general, todos los modelos presentan desafíos en términos de especificidad, como lo reflejan las altas tasas de falsos positivos, lo que subraya la necesidad de mejorar la capacidad de los modelos para distinguir entre clases con mayor precisión.

Recomendaciones

Mejorar la especificidad: Una de las principales limitaciones observadas fue la baja especificidad en todos los modelos, lo que resultó en un alto número de falsos positivos. Para abordar esto, se recomienda:

Implementar técnicas de regularización, como abandono o regularización L2, para mejorar la generalización de los modelos y reducir el sobreajuste.

Ajustar los umbrales de clasificación para encontrar un equilibrio más adecuado entre sensibilidad y especificidad.

Ampliar y equilibrar el conjunto de datos: El desempeño de los modelos podría beneficiar significativamente de un conjunto de datos más amplio y equilibrado, especialmente en la clase benigna. Esto permitiría que los modelos aprendan patrones más representativos y reduzcan la tasa de falsos positivos.

Combinar arquitecturas: Explorar arquitecturas híbridas que combinen las fortalezas de MobileNet y Xception podrían mejorar el desempeño general. Por ejemplo, MobileNet podría usarse como un extractor de características inicial, seguido de un clasificador personalizado optimizado para esta tarea.

Aumentar la robustez del preprocesamiento: Mejorar la calidad de las imágenes mediante

técnicas avanzadas de preprocesamiento, como segmentación precisa de las lesiones y aumento de datos, podría ayudar a los modelos a aprender características más relevantes para la clasificación.

Evaluación adicional con métricas específicas: Incorporar métricas como el valor predictivo positivo (PPV) y el valor predictivo negativo (NPV) puede proporcionar una comprensión más profunda del desempeño de los modelos en un contexto clínico.

Despliegue clínico supervisado: Antes de implementar estos modelos en un entorno clínico, se recomienda integrarlos en flujos de trabajo supervisados por dermatólogos, permitiendo que sirvan como una herramienta de apoyo en lugar de una solución autónoma.

Referencias

- Acevedo, E., Serna, A., y Serna, E. (2017). *Principios y características de las redes neuronales artificiales. Desarrollo e innovación en ingeniería*, 173(), .
- Benedetti, J. (2024, 1). *Estructura y funcionamiento de la piel*.
- Brizuela Cruz, S. (s.f.). *Implementación de la dermatoscopia en la atención primaria para la valoración de patología cutánea en la consulta de Medicina Familiar*.
- Capas de la piel: MedlinePlus enciclopedia médica ilustración*. (s.f.).
<https://medlineplus.gov/spanish/ency/esp/imagepages/8912.htm>
- Forero, L. A. J. (2024, 5). *Día mundial del melanoma 2024 - Cuenta de Alto Costo*.
<https://cuentadealtocosto.org/cancer/dia-mundial-del-melanoma-2024/>
- Hernández, K. M. C. (2024). *Avances Tecnológicos en el Diagnóstico Precoz de Cáncer de Piel. Estudios y Perspectivas Revista Científica y Académica*, 4(2), 1086–1108.
- Huix Pérez, A. (2024). *Clasificación de lesiones cutáneas mediante IA*.
- ISIC — International Skin Imaging Collaboration*. (s.f.).
<https://www.isic-archive.com/>
- Izaurieta, F., y Saavedra, C. (2000). *Redes neuronales artificiales. Departamento de Física, Universidad de Concepción Chile*, (), .
- Jauregui, A. F. (2023, 8). *Qué son y cómo crear una red neuronal convolucional con Keras*.
- López, A. F. P., Álvarez, M. A. V., Caicedo, N. R. A., y Moreira, M. S. N. (2024). *Abordaje integral del melanoma cutáneo en paciente pediátrico: diagnóstico avanzado y estrategias terapéuticas en dermatología. RECIMUNDO*, 8(1), 523–545.
- López, R. F., y Fernández, J. M. F. (2008). *Las redes neuronales artificiales*. Netbiblo.
- Martínez-González, C. (2021). *Transfer Learning en la detección de cáncer de piel*. (Tesis de

Master no publicada).

Martín Tinaquero, D. (2023). *Detección temprana de cáncer de piel mediante clasificador de imágenes basado en Inteligencia Artificial*.

Pertuz, C. M. P. (2022). *Aprendizaje automático y profundo en python*. Ra-Ma Editorial.

Ramírez, I. N. M., Sejin, S. E. F., Maurcia, D. M., Gutiérrez, M. S. V., Ruiz, D. V., Girón, M. A. L.,

... Mendoza, S. C. B. (2024). *Puesta al Día en el Cancer de Piel: Artículo De Revisión*.

Ciencia Latina Revista Científica Multidisciplinar, 8(2), 2192–2204.

Rouhiainen, L. (2018). *Inteligencia artificial*. Madrid: Alienta Editorial, (), 20–21.

Sanabria Codesal, E. (2023). *Las matemáticas del color*.

Team, K. (s.f.). *Keras documentation: Optimizers*.

<https://keras.io/api/optimizers/>

Torres Payoma, F. (2019). *Investigación del Efecto de la Distancia entre Detectores en la Técnica de Retrodispersión de Neutrones Térmicos* (Tesis Doctoral). doi:

10.13140/RG.2.2.26603.26407

Valentina, M. G. Z. (2023). *Desarrollo de un Algoritmo Computacional basado en Deep Learning para la Clasificación de Residuos Orgánicos*.

<https://repository.unad.edu.co/handle/10596/60395>

Vega, M. Á., Mora, L. M. Q., Badilla, M. V. C., y cols. (2020). *Inteligencia artificial y aprendizaje automático en medicina*. *Revista médica sinergia*, 5(8), e557–e557.

Xception Model and Depthwise Separable Convolutions. (2019, 3).