Desarrollo de un modelo automático para el tamizaje de cáncer de piel en imágenes dermatoscópicas evaluado en población colombiana

Paula Andrea Yusunguaira Martínez

Universidad El Bosque

Facultad de Ingeniería

Programa de Bioingeniería

Bogotá, Colombia

2025

Desarrollo de un modelo automático para el tamizaje de cáncer de piel en imágenes dermatoscópicas evaluado en población colombiana

Paula Andrea Yusunguaira Martínez

Trabajo de grado presentado como requisito para optar al título de:

Bioingeniero

Director(a):

Óscar Mauricio Arias Ballén. Ingeniero electrónico.

Línea de Investigación: Salud de sistemas biológicos

Universidad El Bosque

Facultad de Ingeniería

Bogotá, Colombia

2025

Dedico este trabajo a mi familia, especialmente a mis padres por su amor incondicional y su apoyo constante en cada paso de mi vida académica y personal.

Agradecimientos

Agradezco a la Universidad el Bosque, por brindarme las herramientas académicas y el espacio para desarrollar este proyecto. A los docentes Óscar Mauricio Arias Ballén y Karen Torres por su orientación, paciencia y compromiso durante el proceso de desarrollo de este proyecto, así como todos los aportes brindados para concretar un buen trabajo. A mis compañeras Luna Isabella Torres y Mariana Pérez por siempre mostrarme su apoyo y motivarme a pesar de las dificultades que este proyecto trajo consigo. A mis padres Holman Yusunguaira y Angélica Martínez que fueron mi pilar durante toda mi formación, siempre me brindaron un apoyo incondicional además de su amor, apoyo y confianza inquebrantable, que me impulsaron a culminar esta etapa con dedicación y esperanza. A mis familiares Alejandra Buitrago, Eduardo Ruiz y Pedro Martínez quienes además de los conocimientos brindados siempre mostraron su apoyo en la preparación de este proyecto ayudándome a tener la mente enfocada y a fortalecer mi seguridad. Finalmente, a todos aquellos que de una u otra forma hicieron parte de este camino, gracias por acompañarme en esta meta cumplida.

CONTENIDO

Pág.

Resumen 10

Abstract 10

INTRODUCCIÓN 11

1. PLANTEAMIENTO DEL PROBLEMA 13

2. OBJETIVOS 18

3. MARCO REFERENCIAL 19

4. METODOLOGÍA 27

5. RESULTADOS Y DISCUSIÓN 49

6. CONCLUSIONES 75

7. RECOMENDACIONES 76

8. ANEXOS 77

BIBLIOGRAFÍA 83

LISTA DE TABLAS

Pág.

Tabla 1. Técnicas de preprocesamiento de datos …… 25

Tabla 2. Requerimientos generales del modelo ……... 28

Tabla 3. Matriz de decisión ponderada para la elección de arquitecturas para la CNN …. 33

Tabla 4. Relación señal a ruido promedio en técnicas de preprocesamiento…….. 49

Tabla 5. Resultados de balanceo de clases con SMOTE+ENN y SMOTE+Tomek links.. 50

Tabla 6. Combinaciones de hiperparámetros evaluadas en la grilla…… 52

Tabla 7. Reporte inicial de entrenamiento para cada modelo…….. 53

Tabla 8. Valores evaluados en la *grid search* con sus resultados……... 55

Tabla 9. Métricas de rendimiento global de los modelos con Fine-tuning ……. 57

Tabla 10. Reporte de clasificación para cada clase ……... 57

Tabla 11. Métricas globales del modelo en validación con pieles colombianas ……. 66

Tabla 12. Métrica local para fototipo de piel II …….. 68

Tabla 13. Métrica local para fototipo de piel III ……..69

Tabla 14. Métrica local para fototipo de piel IV ……..70

Tabla 15. Métrica local para fototipo de piel V …….. 71

Tabla 16. Métrica local para fototipo de piel VI ……. 72

LISTA DE FIGURAS

Pág.

Figura 1. Datos obtenidos de ISIC Archive ……..... 29
Figura 2. Técnica 1 de Data Transformation en una imagen …….. 30
Figura 3. Técnica 2 de Data Transformation en una imagen …….. 31
Figura 4. Técnica 3 de Data Transformation en una imagen …….. 32
Figura 5. Arquitectura del modelo ResNet50 ……. 40
Figura 6. Arquitectura del modelo MobileNet …… 40
Figura 7. Arquitectura del modelo Inception V3 ……..... 41
Figura 8. Escala de fototipos de piel de Fitzpatrick ……... 43
Figura 9. Boceto encabezado de interfaz ……..……..... 45
Figura 10. Boceto pie de página de interfaz ……..……..... 45
Figura 11. Boceto contenido de la pestaña 'INICIO' 45
Figura 12. Boceto contenido de la pestaña 'ESCANEA' ……..… 45
Figura 13. Boceto contenido de la Pestaña 'ACERCA DEL PROYECTO' ……. 45
Figura 14. Boceto contenido de la pestaña 'INFORMACIÓN ADICIONAL' ……..... 45
Figura 15. Boceto contenido de la pestaña 'TYC' ……. 46
Figura 16. Distribución de los datos con SMOTE+ENN …….….. 52
Figura 17. Curvas de entrenamiento de los modelos ……..... 58
Figura 18. Resultado Ventana 'INICIO' …… 61
Figura 19. Resultado Ventana 'ESCANEA' ……...…... 62
Figura 20. Resultado del informe de predicción del modelo con una imagen de prueba… 62
Figura 21. Resultado Ventana 'ACERCA DEL PROYECTO' ……..……..…... 63
Figura 22. Resultado Menú Melanoma ……..……..... 64
Figura 23. Resultado Menú CEC ……...……. 64
Figura 24. Resultado Menú CBC …… 64
Figura 25. Resultado Menú Lesiones Benignas …….. 64
Figura 26. Resultado Ventana 'T & C' ……..……. 65

Figura 27. Matriz de confusión resultado global……. 66
Figura 28. Matriz de confusión para fototipo de piel II ……...…... 68
Figura 29. Matriz de confusión para fototipo de piel III ……..…... 69
Figura 30. Matriz de confusión para fototipo de piel IV ……..…... 70
Figura 31. Matriz de confusión para fototipo de piel V ……..... 71
Figura 32. Matriz de confusión para fototipo de piel VI ……. 72

LISTA DE ANEXOS

Pág.

Anexo A. Diagrama de flujo del modelo …… 77

Anexo B. Diagrama de flujo de navegación ……... 78

Anexo C. Manual de usuario …….. 79

Resumen

El cáncer de piel representa una de las principales causas de morbilidad oncológica a nivel mundial, y su diagnóstico temprano resulta esencial para reducir la mortalidad asociada. Este trabajo tuvo como objetivo desarrollar y evaluar un modelo automático basado en redes neuronales convolucionales (CNN) para la clasificación de lesiones cutáneas malignas y benignas a partir de imágenes dermatoscópicas, con un enfoque de validación en población colombiana. El modelo se entrenó utilizando la base de datos pública *ISIC Archive* y se probó con un conjunto de imágenes representativas de distintos fototipos de piel de Colombia según la escala de Fitzpatrick. Se calcularon métricas de desempeño global y por tipo de piel para interpretar las predicciones. Los resultados de implementación inicial mostraron una precisión superior al 85 % y una disminución al 72 % al validarlo con las imágenes colombianas, indicando la necesidad de mejorar la representatividad de ciertos fototipos en los datos. Se concluye que la capacidad predictiva se ve influenciada por la representatividad demográfica en el entrenamiento de modelos de aprendizaje, por lo que resulta esencial disponer de conjuntos de datos locales diversos que reflejen las características de la población colombiana. Los hallazgos confirman el potencial del aprendizaje profundo como herramienta de apoyo en el tamizaje de cáncer de piel, aunque requieren adaptaciones locales para lograr modelos equitativos y clínicamente confiables.

Palabras clave: cáncer de piel; redes neuronales convolucionales; dermatoscopia; fototipo cutáneo; población colombiana.

Abstract

Skin cancer is one of the leading causes of cancer morbidity worldwide, and early diagnosis is essential to reduce associated mortality. The aim of this study was to develop and evaluate an automatic model based on convolutional neural networks (CNN) for the classification of malignant and benign skin lesions from dermoscopic images, with a focus on validation in the Colombian population. The model was trained using the public ISIC Archive database and evaluated with a set of images representative of different Colombian skin phototypes according to the Fitzpatrick scale. Overall performance metrics and phototype skin metrics were calculated to interpret predictions. The primary implementation results showed an accuracy of over 85% and a decrease to 72% when validated with Colombian images, indicating the need to improve the representativeness of certain phototypes in the data. It is concluded that predictive capacity is influenced by demographic representativeness in the training of learning models, making it essential to have diverse local datasets that reflect the characteristics of the Colombian population. The findings confirm the potential of deep learning as a support tool in skin cancer screening, although local adaptations are required to achieve equitable and clinically reliable models.

Key words: skin cancer; convolutional neural networks; deep learning; dermoscopy skin phototype; Colombian population.

INTRODUCCIÓN

El cáncer de piel se encuentra entre las diez neoplasias más comunes a nivel mundial, con predominio de los carcinomas basocelulares (75 %) y escamo celulares (20 %), mientras que el melanoma, aunque representa alrededor del 5% de los casos, es el tipo más agresivo por su alta capacidad de invasión y metástasis, siendo responsable de aproximadamente el 75 % de las muertes asociadas a esta enfermedad [Ospina et al., 2021; World Health Organization, 2023]. La carga epidemiológica del cáncer de piel constituye una preocupación creciente para la salud pública global, especialmente por su impacto en la mortalidad evitable y en la calidad de vida de los pacientes. En Colombia, las tendencias muestran un incremento sostenido en su incidencia, particularmente en poblaciones expuestas a radiación solar de manera continua, como las comunidades rurales del sector agropecuario [Instituto Nacional de Cancerología, 2022]. Esta situación subraya la urgencia de fortalecer las estrategias de diagnóstico temprano y tamizaje oportuno, con el fin de reducir la morbilidad y mortalidad asociadas.

Tradicionalmente, el diagnóstico del cáncer de piel ha dependido de la observación clínica directa por parte del médico, complementada ocasionalmente con técnicas como la dermatoscopia. Aunque esta herramienta ha demostrado mejorar la sensibilidad diagnóstica, su efectividad sigue estando condicionada por la experticia del profesional, lo que introduce una importante subjetividad en el proceso. Además, los métodos de confirmación diagnóstica como la biopsia, aunque precisos, son invasivos, costosos y generan molestias al paciente, lo que retrasa el inicio del tratamiento y disminuye las tasas de supervivencia [Nogueira et al., 2021].

En este contexto, la inteligencia artificial (IA) y, en particular, las redes neuronales convolucionales (CNN), han demostrado un gran potencial en el análisis automatizado de imágenes médicas, proporcionando diagnósticos rápidos, reproducibles y no invasivos [Brinker et al., 2019]. Pues a nivel global, diversos estudios han implementado modelos computacionales de clasificación de lesiones cutáneas utilizando estas técnicas, mostrando resultados prometedores en términos de precisión diagnóstica. Sin embargo, la mayoría de los modelos desarrollados y validados en este campo utilizan bases de datos provenientes de poblaciones europeas o norteamericanas, cuyos fototipos cutáneos y características dermatológicas difieren significativamente de la población colombiana, este desajuste puede afectar la precisión diagnóstica y generar sesgos al aplicar estos modelos en nuestro contexto. Por tanto, surge la necesidad de evaluar y adaptar estas tecnologías a nuestro contexto específico, validando su desempeño sobre imágenes de pacientes colombianos.

Frente a esta brecha, el presente trabajo de grado tiene como objetivo principal el desarrollo de un modelo automático para el tamizaje de cáncer de piel en imágenes dermatoscópicas, entrenando inicialmente bases de datos internacionales y posteriormente evaluado con

imágenes provenientes de población colombiana. El propósito de esta validación es identificar la capacidad del modelo para adaptarse a los fototipos de piel predominantes en el país, garantizando así su fiabilidad y eficacia en el entorno clínico nacional. Este enfoque responde a una necesidad práctica: disminuir los tiempos de diagnóstico, mejorar la cobertura en zonas de difícil acceso a dermatólogos especializados, y reducir el impacto económico y emocional que genera el cáncer de piel cuando es detectado en fases avanzadas.

La metodología de este estudio se estructuró siguiendo un enfoque secuencial que abarcó desde la obtención y preparación de las imágenes dermatoscópicas hasta la construcción, entrenamiento y validación de un modelo automático para el tamizaje de cáncer de piel. En términos de alcance, este proyecto pretende sentar bases para la integración de herramientas automáticas de diagnóstico en el sistema de salud colombiano, lo cual representaría una solución de bajo costo y amplia cobertura para zonas rurales o apartadas, donde el acceso a especialistas es limitado. No obstante, también se reconocen las limitaciones del estudio: el uso de imágenes internacionales para el entrenamiento inicial puede afectar la representatividad del modelo, y el tamaño del conjunto local de validación, aunque relevante, aún puede ser insuficiente para afirmar su eficacia generalizada sin una implementación piloto en campo clínico real.

Como contribución científica, este trabajo representa un avance significativo en la bioingeniería aplicada a la salud pública colombiana porque propone un modelo de atención, que aprovecha las capacidades del aprendizaje automático para ofrecer diagnósticos rápidos, no invasivos y más accesibles, especialmente para poblaciones vulnerables. La validación en población local es un aspecto que brinda información sobre cómo cambia la capacidad predictiva de un modelo de tamizaje de cáncer de piel cuando se enfrenta a un grupo de validación con fototipos de piel más diversos poniendo de manifiesto la importancia de implementar tecnologías accesibles y confiables que, apoyadas por profesionales del área, mejoren significativamente los resultados clínicos y la calidad de vida de los pacientes.

1. PLANTEAMIENTO DEL PROBLEMA

1.1 DEFINICIÓN DEL PROBLEMA

El cáncer es un término genérico que se utiliza para designar un amplio grupo de enfermedades que pueden afectar cualquier parte del organismo [World Health Organization: WHO, 2022], estas afecciones se desarrollan debido a la rápida multiplicación de células anormales que se extienden por fuera de sus límites invadiendo partes adyacentes del cuerpo o incluso propagándose a otros órganos causando, en la mayoría de los casos, la muerte del paciente. Actualmente, el cáncer es una enfermedad letal y la principal causa de muerte a nivel mundial, con alrededor de 10 millones de defunciones anuales atribuidas a esta patología [World Health Organization: WHO, 2022]. Según la Organización Mundial de la Salud, desde 2020 los cánceres más comunes debido a su incremento en los últimos años son: cáncer de mama, con 2.26 millones de casos; cáncer de pulmón, con 2.21 millones; cáncer colorrectal, con 1.93 millones; cáncer de próstata, con 1.41 millones; cáncer de piel, con 1.20 millones; y cáncer gástrico, con 1.09 millones.

El cáncer de piel puede presentarse de tres formas principalmente: carcinoma basocelular, que corresponde al 75% de los casos; carcinoma escamo celular, que representa el 20%; y melanoma, que abarca el 5%. A pesar de ser el menos frecuente, el melanoma es el más agresivo, ya que tiene mayor probabilidad de invadir tejidos cercanos y diseminarse a otras partes del cuerpo. Debido a estas características, el melanoma es responsable del 75% de las muertes por cáncer de piel [Ospina et al., 2021].

El melanoma es un tipo de cáncer cuyo diagnóstico temprano es fundamental para la supervivencia del paciente, ya que permite aplicar un tratamiento oportuno y aumentar las posibilidades de una cura definitiva. Sin embargo, la detección temprana de melanoma es un desafío para los médicos, se estima que el 50% de los pacientes se diagnostican en estadios avanzados y el 55% en estados metastásicos [Serrano & Barrera, 2019], ocasionando que el 30% de los fallecimientos se dé por una detección tardía de la enfermedad [Secretaría de Salud, 2024]. En el caso del cáncer no melanoma, éste es menos agresivo pues su crecimiento es más lento y su potencial de metástasis es muy bajo (0.0020%), sin embargo, tienen una importante capacidad de invasión y destrucción local de tejidos adyacentes lo que puede desencadenar alteraciones funcionales importantes [Instituto Nacional de Cancerología Colombia, 2021], esto ocasiona que más de 5400 personas en todo el mundo mueran a causa de estos padecimientos [The Skin Cancer Foundation, 2024].

En Colombia, como en la mayoría de los países, el proceso de atención para pacientes con sospecha de cáncer de piel comienza con la evaluación de un médico general, quien determina si es necesario remitir al paciente a un dermatólogo. El dermatólogo, tras realizar una anamnesis y una inspección visual, decide si es pertinente realizar una biopsia para confirmar el diagnóstico. Actualmente, en los entornos de atención primaria, se está

implementando la dermatoscopia, una técnica que utiliza aumento y luz visible para examinar la piel con mayor detalle en comparación con la observación a simple vista, esta herramienta ha mejorado la sensibilidad del diagnóstico clínico, situándose entre el 60% y el 90%, dependiendo en gran medida de la experticia del profesional que la emplea [Dinnes et al., 2018].

Otro factor que dificulta el diagnóstico temprano de cáncer de piel es distinguir entre lesiones benignas (Lunares, Queratosis seborreica, Hemangiomas, Lipomas o Verrugas) y la enfermedad en su etapa inicial, especialmente un tipo de lunar llamado 'nevo Spitz' que incluso si se observa con un microscopio puede ser difícil de identificar, para verificar que se trata de un melanoma se debe hacer una biopsia, un proceso invasivo que se puede realizar por raspado, punción, escisión o incisión [American Cancer Society, 2019], en la mayoría de casos el resultado de este procedimiento tarda entre tres a cinco días y tiende a ser incómodo para el paciente pues este puede generar pequeños moretones, sangrado o incluso hasta dolor. El funcionamiento de esta línea de atención médica y dado que los métodos de diagnóstico actuales están sujetos a la dependencia del criterio del médico, el acceso limitado a especialistas y la subjetividad de las evaluaciones; ha ocasionado que el tiempo de espera para recibir un diagnóstico de cáncer de piel en Colombia sea de 48 días y para recibir el tratamiento correspondiente de casi 83 días [Barrera, 2022].

A pesar de los avances en técnicas como la dermatoscopia, el diagnóstico temprano de cáncer de piel sigue siendo altamente dependiente de la experiencia del médico y como ya se mencionó anteriormente, requiere de procedimientos invasivos como la biopsia, este es un enfoque muy tradicional, que además se ve limitado por diferentes factores como los tiempos de espera y la subjetividad inherente a la interpretación visual de las lesiones cutáneas. Según esto, hay una evidente brecha tecnológica considerando la gran variedad de oportunidades que ofrece la tecnología para el diagnóstico de enfermedades actualmente, pues se ha evidenciado como el uso de herramientas automatizadas o basadas en inteligencia artificial pueden asistir a los profesionales de salud en la identificación precisa y rápida de variedad de patologías utilizando técnicas de procesamiento de imágenes.

Actualmente, la mayoría de los avances en el desarrollo de modelos de detección basados en técnicas de procesamiento de imágenes se han realizado en Estados Unidos y Europa, donde todos los estudios se han desarrollado en torno a sus poblaciones y sus respectivos fototipos de piel, la mayoría de estos modelos no solo reducen significativamente los tiempos de diagnóstico, sino que también aumentan la precisión del mismo, disminuyendo la dependencia del criterio subjetivo y facilitando el acceso a diagnósticos de alta calidad. Sin embargo, en Colombia estas herramientas para uso clínico generalizado enfocadas en la población colombiana y sus características dermatológicas, no son muy populares en las líneas de atención primaria del sistema de salud; esto limita la capacidad de los médicos para ofrecer diagnósticos rápidos y efectivos, lo que acentúa la demora en el tratamiento y reduce las tasas de detección temprana.

1.2 JUSTIFICACIÓN

En las últimas décadas, la incidencia del cáncer de piel ha aumentado considerablemente a nivel mundial, con más de 1.5 millones de casos nuevos reportados [Collazos, 2021]. En Colombia, se estima que la tasa de incidencia es de aproximadamente 102 casos nuevos por cada 100,000 personas, lo que representa alrededor de 26,456 casos anuales [Velásquez, 2020]. Como se ha mencionado anteriormente el melanoma es la forma más agresiva de esta neoplasia y su diagnóstico temprano es crucial para mejorar las tasas de supervivencia. Actualmente es catalogado como la quinta causa de muerte por cáncer en el mundo y se le atribuyen el 70% de las muertes producidas por cáncer de piel, en Colombia se estima que la incidencia de melanoma maligno en mujeres es de 4.6 por cada 100.000 y en hombres 4.4 por cada 100.000, estas cifras de incidencia se han incrementado en un 3% desde el 2020 y continua con esa tendencia creciente [Uribe et al, 2021]. Entre la población colombiana esta afección tiene una frecuencia del 16.1% y actualmente hay alrededor de 7.401 personas diagnosticadas [Uribe et al, 2021]; entre estos, según el Dr. Santiago Andrés Ariza, presidente de la asociación colombiana de dermatología oncológica (ACDOC), el 80% de los casos de cáncer de piel del país se dan en el sector agropecuario, además el 50% se dan en personas que oscilan entre los 42 a 91 años y 69% de esta cifra corresponde a mujeres [Velasquez, 2020].

El riesgo de desarrollar cáncer de piel también está influenciado por el tipo de piel de la población. En Colombia, aunque existe diversidad en los tipos de piel, predomina la piel trigueña o morena, también conocida como piel latina. Este tipo de piel presenta un riesgo de 0.6%, lo que equivale a 1 de cada 167 personas, siendo el segundo grupo con mayor probabilidad de desarrollar cáncer de piel, después de la piel blanca [Uribe et al., 2021].
La agresividad del cáncer de piel varía según la fase de desarrollo en la que se encuentra y el tipo de lesión, en su etapa 'in situ', el cáncer se limita a la epidermis sin invadir los tejidos más profundos, pero si este evoluciona a invasor, adquiere la capacidad de invadir el tejido conectivo subyacente, formando tumores con un grosor mucho más pronunciado y estos son los que tienen un mayor riesgo de metastatizar. En su fase metastásica, el cáncer puede diseminarse a los ganglios linfáticos y órganos distantes como los pulmones, el hígado, el cerebro y los huesos [Casariego & Baudo, 2004]. Las metástasis del cáncer de piel pueden tener consecuencias devastadoras para el organismo. A nivel pulmonar, puede causar tos persistente, dificultad para respirar y, en algunos casos, hemoptisis; en el cerebro, los daños pueden variar desde dolores de cabeza hasta convulsiones, alteraciones en la visión y cambios en el comportamiento; el hígado, al verse afectado, altera su función metabólica, mientras que en los huesos, la estructura ósea se debilita, aumentando el riesgo de fracturas; asimismo, la afectación de los ganglios linfáticos compromete el sistema inmunológico [Cancer Center MOFFITT, 2023]. Estas complicaciones subrayan el peligro que representa esta patología.

Otro factor relevante es el económico, ya que los pacientes deben asumir los altos costos asociados al tratamiento de la enfermedad. En los casos avanzados, los procedimientos son

más complejos y el costo del tratamiento puede oscilar entre $1,000 y $5,000 dólares, dependiendo del tamaño y la extensión de los tumores [Skin Cancer Financial Assistance, 2023]. En contraste, tratar el cáncer de piel en sus etapas más tempranas puede ser significativamente más accesible, por ejemplo, la terapia fotodinámica, utilizada en fases iniciales, tiene un costo entre $200 y $600 dólares, la quimioterapia varía entre $300 y $600 dólares, y la crioterapia se encuentra por debajo de los $1,000 dólares [Skin Cancer Financial Assistance, 2023]. Estos costos adquieren mayor relevancia cuando se considera el contexto de quienes padecen la enfermedad, ya que el 80% de los casos en Colombia se presentan en el sector agropecuario, afectando principalmente a campesinos y personas con exposición constante al sol, esto hace que, para ellos sea aún más difícil acceder a los recursos necesarios para cubrir los tratamientos. Estas cifras subrayan la importancia de un diagnóstico temprano, no solo para mejorar las tasas de supervivencia, sino también para reducir el impacto financiero sobre los pacientes.

El bioingeniero puede desempeñar un papel crucial en la mejora de los diagnósticos de cáncer de piel mediante el desarrollo de tecnologías avanzadas que combinan inteligencia artificial (IA) y procesamiento de imágenes dermatoscópicas como complemento en el proceso de diagnóstico, ayudando a reducir el tiempo necesario para identificar las lesiones y facilitando la toma de decisiones sobre biopsias o tratamientos.

La IA, a través del uso de redes neuronales convolucionales (CNN) y técnicas de *machine learning*, permite la creación de modelos que pueden analizar grandes volúmenes de imágenes con una precisión superior a la de los métodos tradicionales [Juan R.Q. et al, 2011]. Estos modelos tienen el potencial de aprender a identificar patrones específicos de lesiones cutáneas, diferenciando entre lesiones benignas y malignas, y detectando características que podrían pasar inadvertidas en una inspección visual.

Otro punto importante es que la bioingeniería puede ofrecer soluciones accesibles y portátiles, como aplicaciones móviles o dispositivos conectados a cámaras de dermatoscopia, que puedan ser utilizados por médicos generales o en áreas rurales con acceso limitado a especialistas; esto podría ampliar la capacidad de diagnóstico, permitiendo que más pacientes reciban un análisis rápido y confiable sin la necesidad de recurrir a biopsias invasivas en fases tempranas.

Según esto la propuesta de este proyecto es desarrollar un modelo automático de tamizaje para detección temprana de cáncer de piel y evaluar su implementación en la población colombiana utilizando los fundamentos del procesamiento de imágenes y la inteligencia artificial. Aunque los modelos existentes han sido entrenados utilizando bases de datos internacionales, es objetivo es evaluar su desempeño en un contexto local debido a las variaciones en las características dermatológicas de los pacientes colombianos, principalmente en los fototipos de piel, las cuales pueden influir en la precisión del modelo, ya que los algoritmos entrenados en datos de otras regiones pueden llegar a presentar sesgos al aplicar sus resultados en población local. Por ello, este proyecto se centra en validar la

efectividad del modelo con imágenes provenientes de la población colombiana, con el fin de determinar su capacidad para adaptarse a las características particulares de la piel en este contexto y garantizar que el modelo sea confiable para el diagnóstico en el país.

A futuro ese modelo se podría integrar con una plataforma digital o una aplicación móvil para permitirle al usuario tomar una imagen y obtener un análisis preliminar en cuestión de segundos. Este proyecto no solo mejoraría la precisión del diagnóstico, sino que también reduciría la dependencia de la interpretación subjetiva del médico y aceleraría el proceso de referencia a un especialista cuando sea necesario. Se espera que este proyecto contribuya a la modernización del sistema de salud colombiano, aprovechando la bioingeniería para mejorar las tasas de supervivencia mediante la detección temprana de cáncer de piel, impactando positivamente la salud pública y la eficiencia de los servicios médicos en el país.

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Desarrollar un modelo automático para el tamizaje de cáncer de piel en imágenes dermatoscópicas de lesiones cutáneas y evaluarlo en imágenes de población colombiana

2.2. OBJETIVOS ESPECÍFICOS

Diseñar un modelo de detección automática para el tamizaje de cáncer de piel en imágenes dermatoscópicas de lesiones cutáneas, utilizando bases de datos internacionales.

Implementar el modelo de detección automática para el tamizaje de cáncer de piel en imágenes dermatoscópicas de lesiones cutáneas, utilizando bases de datos internacionales.

Validar el modelo de detección automática para el tamizaje de cáncer de piel en imágenes dermatoscópicas con imágenes de la población colombiana, evaluando la viabilidad del modelo en fototipos de piel específicos de esta población.

3. MARCO REFERENCIAL

En esta sección se encuentra la revisión del estado del arte y el desarrollo de los conceptos relacionados.

3.1. ESTADO DEL ARTE

En esta revisión se exploraron los avances recientes en el campo de la detección de cáncer de piel a través de variedad de técnicas de procesamiento de imágenes, identificando las principales limitaciones en la implementación de estos sistemas en diferentes contextos. Los primeros hallazgos están focalizados en la detección de melanoma, sin embargo, son de gran utilidad pues estas técnicas se pueden extrapolar a todos los tipos de cáncer de piel.

Un artículo publicado en Barcelona implementó una técnica diferente para la detección de lesiones cutáneas malignas, el *Deep Learning*, este ámbito de la inteligencia artificial permite extraer características de conjuntos de datos manejando una estructura lógica por capas; en este artículo se tiene como objetivo desarrollar una red neuronal predictiva basada en esta técnica con el fin de identificar lesiones dermatológicas que requieran atención especializada; para ello se utilizó un modelo pre entrenado (*Transfer Learning*) con el que se buscaba transferir el reconocimiento de imágenes de un dataset genérico a la clasificación del estudio en particular; pasos adicionales vistos en la metodología son el balanceo de la data y la optimización de hiperparámetros y variables, finalmente se logró llegar a una precisión del 84% y uno de los principales factores que se tuvieron que tomar en cuenta para llegar a ese porcentaje fueron la resolución de las fotos y el overfitting [Espasa,2022]; finalmente estudiaron la aplicación del modelo en escenarios reales y su potencial comercialización.

Otro referente dentro de esta línea de tiempo es el artículo publicado en 2023 en Buenos Aires, Argentina, titulado "Aplicación de redes neuronales densas y convolucionales para detección de COVID-19 en imágenes de rayos X". Su objetivo principal fue determinar qué arquitecturas de Redes Neuronales Convolucionales (CNN) y Redes Neuronales Densas lograban un mejor desempeño en la identificación de casos de neumonía y COVID-19 a partir de imágenes de tórax. Para ello, los autores utilizaron un conjunto de 603 imágenes provenientes de bases de datos públicas, aplicaron una validación Hold-out y desarrollaron una CNN de 13 capas que obtuvo una precisión del 94.73%. Esta propuesta surge como una adaptación de una MLP (Multi-Layer Perceptron) de 12 capas y de una ResNet de 18 capas [Cruz R.S. & Delrieux C.C., 2023]. El estudio constituye un aporte relevante, ya que evidencia la versatilidad de las CNN como herramientas de apoyo diagnóstico basadas en imágenes médicas, demostrando su capacidad para aprender patrones distintivos y realizar clasificaciones precisas en conjuntos de datos heterogéneos.

Otro artículo publicado en la India, titulado 'Detección y clasificación del cáncer de piel melanoma mediante la técnica de procesamiento de imágenes' y publicado en 2023 propone métodos semejantes al artículo mencionado anteriormente, a diferencia que en este se

proponen dos opciones de detección asistida por ordenador para el diagnóstico precoz del melanoma con el fin de comparar cuál de los obtienen un mayor precisión y puede ser implementado como una aplicación web; el primer método emplea redes neuronales convolucionales, con esta técnica se obtuvo una tasa de precisión del 94% después de 100 épocas de cómputo; por otro lado, el segundo método propuesto utilizó máquinas de soporte vectorial SVM con un kernel RBF y con tres clases (benignas, malignas o normales) y obtuvo una precisión del 86.6% [Viknesh et al, 2023]. Los autores recalcan que la alta precisión obtenida es a razón de que, en lugar de utilizar algoritmos tradicionales de IA para la detección de enfermedades de la piel, tomaron un enfoque basado en el aprendizaje profundo para automatizar la extracción de características; sin embargo, una de las principales desventajas de este sistema es el tiempo de respuesta del modelo y el tamaño del lote que probablemente se podrían corregir con conjuntos de datos más grandes [Viknesh et al, 2023].

Ya enfocándonos en artículos mucho más recientes, se encontró 'Clasificación de lesiones cutáneas mediante IA' publicado en Alicante en mayo del presente año donde se busca desarrollar un modelo computacional capaz de procesar y clasificar imágenes dermatoscópicas, para ello utilizaron el dataset de ISIC que tiene 25.331 imágenes de diferentes categorías diagnósticas la cual tuvo que ser procesada para homogeneizar los metadatos de los subconjuntos de entrenamiento y prueba; así mismo estas imágenes fueron preprocesadas ajustando su tamaño, transformaciones de aumento de datos, normalización, etiquetado y balanceo de clases; para el clasificador utilizaron la arquitectura de red neuronal convolucional ResNet-34 que se caracteriza por tener en sus capas un kernel de 7x7 y una capa de max pooling; adicionalmente utilizaron una función de pérdida de entropía cruzada ponderada [Huix, 2024]. Con estas características se construyó, entrenó y validó el modelo, donde se obtuvo una precisión del 93,7% con altos valores en las métricas del modelo y gran equilibrio entre sensibilidad y especificidad; a pesar de los buenos resultados mostrados dos de las clases (melanoma y carcinoma de células escamosas) fueron el principal desafío por su menor representación y mayor dificultad diagnóstica [Huix, 2024].

Otro aporte relevante dentro del desarrollo reciente de herramientas basadas en aprendizaje profundo para el diagnóstico dermatológico es el estudio realizado en México sobre la detección automática de carcinoma basocelular mediante redes neuronales convolucionales (CNN). En este trabajo se contextualiza la importancia clínica del problema: los cánceres cutáneos representan el segundo tipo de cáncer más frecuente en el país (13%), y el carcinoma basocelular constituye entre el 60% y el 73% de las neoplasias malignas de la piel, apareciendo en aproximadamente el 10% de las biopsias cutáneas, con mayor prevalencia en adultos mayores de 50 años. Considerando esta incidencia, el estudio propone el uso de CNN para la identificación de carcinoma basocelular, especialmente en la región facial, donde esta lesión aparece con mayor frecuencia. Para ello, emplearon bases de datos con imágenes dermatoscópicas y convencionales de lesiones cutáneas, las cuales fueron sometidas a estandarización en formato y tamaño antes del entrenamiento del modelo. El enfoque permitió que la red neuronal aprendiera características específicas de este tipo de cáncer, alcanzando una precisión cercana al 97.6% en la detección de carcinoma basocelular en

pacientes adultos [Morales, A.H., Lepe, D.R. & Díaz, J.F.; 2024]. Este trabajo evidencia la efectividad del aprendizaje profundo como herramienta complementaria para la evaluación clínica y resalta su potencial en el apoyo temprano al diagnóstico en poblaciones de alto riesgo.

Otras metodologías que implican el uso de CNN, donde no se diseña una arquitectura específica sino que se evalúan en simultaneo varias arquitecturas y se comparan sus resultados para ver cuál es la más funcional en un contexto específico, se muestra en el articulo 'Redes Neuronales Convolucionales para detección de retinopatía diabética' de 2025, allí se evaluaron comparativamente ResNet-18, ResNet-50 y una CNN personalizada no preentrenada, la cual alcanzó una precisión del 91% superando a los otros modelos implementados, en este proyecto la principal conclusión radicó en que a complejidad no garantiza necesariamente un mejor desempeño, y que un diseño cuidadoso puede superar modelos preentrenados en tareas de diagnóstico por imagen cuando la cantidad de datos con que se cuenta es limitada [Patino Pérez D. et al. 2025].

Esta revisión documental ha permitido evidenciar los significativos avances en la clasificación de lesiones cutáneas utilizando variedad de técnicas de procesamiento de imágenes recalcando el uso de la inteligencia artificial y *Deep Learning*; a pesar de esto aún persisten muchos desafíos y brechas asociadas a las limitaciones de la tecnología y las complejidades intrínsecas de la dermatología y la imagenología médica.

3.2. MARCO TEÓRICO Y CONCEPTUAL

3.2.1. Cáncer de piel y lesiones cutáneas

El cáncer de piel es una de las neoplasias más comunes en el mundo, este se desarrolla en la epidermis, la capa externa de la piel y que está compuesta por tres tipos de células: las escamosas que son delgadas y planas y forman la capa superior de la epidermis; las células basales son redondeadas y se ubican por debajo de las escamosas; finalmente los melanocitos, ubicados en la parte inferior de la epidermis y productoras de melanina, el pigmento de que le da color a la piel [Collazos, 2021].

Esta enfermedad se puede clasificar en dos tipos principales: Cáncer de piel no melanoma (donde se encuentran se encuentran el carcinoma basocelular y el carcinoma escamo celular o de células espinosas) y melanoma.

Carcinoma basocelular (CBC):

Es la neoplasia maligna más frecuente y suele afectar a las personas de edad avanzada, el tumor deriva de las células que forman la capa basal de la epidermis, pues estas comienzan a crecer hacia abajo y se muestran empanizadas en la periferia de un lóbulo de células tumorales comenzando a formar estos tumores que se caracterizan por su crecimiento lento,

invasión local y no metástasis; visualmente se manifiesta como una lesión perlada y puede sangrar o desarrollar una costra [Soto et al, 2018].

Uno de los factores de riesgo más significativo de este tipo de cáncer es la inmunosupresión inducida por las radiaciones UV, pues esto favorece algunos mecanismos inmunológicos supresores que alteran la función de presentación de antígenos de las células de Langerhans, lo cual permite el flujo de células inflamatorias [Vidrio, 2003]. De igual manera, otros factores como: genéticos (mutación del gen supresor P-53), substancias carcinógenas, los rayos X, infecciones o úlceras crónicas [Vidrio, 2003].

Dentro del cuadro clínico se contemplan cuatro tipos fundamentales de lesiones:

- Exofíticas: forma tumoral que aparece como una lesión hemisférica, eritematosa o violácea.
- Planas: forma superficial constituida por placas eritematosas, a nivel físico su principal característica es el borde brillante.
- Ulceradas: forma terebrante que se caracteriza por lesiones ulceradas desde el comienzo, presenta infiltración y destrucción de tejidos vecinos.
- Pigmentadas: cuando la pigmentación es suficiente para constituir una parte importante del tumor [Friedman et al, 1991].

Su diagnóstico se centra en el análisis de su morfología característica y se confirma mediante biopsia o histopatología; para su tratamiento se puede optar por procedimientos tanto quirúrgicos como no quirúrgicos. El riesgo de metástasis es muy bajo y la mayoría de los tumores son curables.

Carcinoma de células escamosas (CSC):

Este tipo de cáncer de piel se desarrolla en las células escamosas de la epidermis o sus anexos usualmente en zonas que han estado expuestas frecuentemente al sol, es menos común que CBC, pero más propenso a invadir ganglios regionales u otros órganos (hacer metástasis), físicamente es más duro al tacto y se siente escamoso [Collazos, 2021]. Esta tiene un crecimiento rápido y suele aparecer sobre lesiones precancerosas como las queratosis actínicas o úlceras crónicas. Predomina en la cara, pero a su vez se puede encontrar en extremidades, genitales y mucosas orales y anales [Vidrio, 2003].

Los factores de riesgo asociados son: hidrocarburos, radiaciones ionizantes o ultravioleta, enfermedades virales, inmunosupresión, exposición solar, piel blanca, ojos claros, edad o tabaquismo.

En general el pronóstico es bueno siempre y cuando sea diagnosticado en su etapa temprana, por otro lado, la metástasis de CEC está asociada a factores como el tamaño del tumor mayor a 2cm, velocidad de crecimiento, ubicación anatómica (más agresivos en mucosas) y diferenciación histológica.

Melanoma:

Neoplasia que se origina en los melanocitos, también se puede describir como la transformación maligna del melanocito; este tumor con gran capacidad de invasión puede

desarrollarse a partir de un lunar existente o aparecer como una nueva mancha oscura en la piel [Acosta et al, 2009], puede aparecer en mucosas, globo ocular, leptomeninges y tracto intestinal y tiene gran capacidad de metastatizar. El melanoma se produce como resultado de complejas interacciones de diferentes factores genéticos (mutación del oncogén N-ras del P-53, gen supresor CDKN2A) y ambientales como la exposición a sustancias (químicas, virus o radiaciones ionizantes) o la luz ultravioleta [Vidrio, 2003].

Algunos de los factores de riesgo más relevantes son: piel blanca, pelo rubio/pelirrojo, ojos claros, bronceado, aumento del número de lunares, inmunosupresión, historia familiar, bronceados o tendencia a quemaduras solares.

Se identifican cuatro tipos básicos de melanoma:

- Melanoma léntigo maligno: se observa en cara o cuello, es el menos agresivo dado que se mantiene *in situ* por varios años hasta induración o ulcera [Vidrio, 2003].
- Melanoma de extensión superficial: inicia como una lesión plana pigmentada que conforme avanza puede llegar a mostrar zonas infiltradas o elevadas [Vidrio, 2003].
- Melanoma nodular: tumor saliente de color casi negro o azul, tiene crecimiento vertical, es invasor con mucha tendencia a diseminarse [Vidrio, 2003].
- Melanoma acral lentiginoso: comienza como una lesión macular que después se hace infiltrada, queratósica con lesiones elevadas y se encuentra en la región plantar de manos y pies. En Latinoamérica es una de las más comunes [Vidrio, 2003].

Al ser tan agresivo el factor más importante para el éxito del tratamiento es el diagnóstico temprano; para este se utiliza la técnica ABCDE (Asimetría, Bordes irregulares, Color, Diámetro, Elevación y Evolución) [Acosta et al, 2009]. Adicionalmente el pronóstico del paciente se define de diferentes maneras posibles: según el nivel de invasión creado por Clark que identifica cinco grados, el método Breslow para medir el espesor del melanoma y las clasificaciones AJC o TNM.

En el contexto de las lesiones cutáneas, se pueden encontrar otro tipo que corresponde al grupo de lesiones benignas, no cancerosas o precancerosas, entre las que se encuentran:

Queratosis actínica: Son lesiones precancerosas que sin tratamiento pueden desarrollarse en carcinoma de células escamosas, se duele desarrollar en áreas de la piel expuestas al sol constantemente, es decir el rostro, las manos y los brazos; aparece como resultado de una proliferación y diferenciación anómala de queratinocitos epidérmicos, se manifiesta como máculas o pápulas eritematosas, que progresivamente se recubren de una escama adherente que al desprenderse deja una erosión superficial [Méndez, 2016].

Nevus melanocitos: Son acumulaciones benignas de melanocitos en la piel, usualmente se caracterizan por su uniformidad en color y forma; en la mayoría de los casos no representan un riesgo para el paciente, sin embargo, pueden evolucionar a melanoma por factores genéticos o ambientales [Collazos, 2021].

Dermatofibroma: es una lesión cutánea benigna de tejido blandos caracterizada por nódulos subcutáneos firmes. Se encuentran principalmente en extremidades y en todos los grupos de edad, es una lesión asintomática que únicamente requiere tratamiento cuando presenta

síntomas como dolor, sensibilidad o picazón. Los tipos más comunes son benigno celular, epitelioide, aneurismático y hemosiderótico [Myers y Filman, 2024].

Queratosis seborreica: es un tipo de lesiones cutáneas benignas y frecuentes, que se presentan en ambos sexos generalmente en la tercera década de la vida, se suele manifestar como maculas, pápulas o placas hiperqueratosicas bien delimitadas, pigmentadas o no, pudiendo ser endofíticas o Exofíticas. Son asintomáticas y tiene varias formas clínicas: queratosis seborreica común, dermatosis papulosa nigra, pediculada, plana y en estuco [Méndez M., 2017].

3.2.2. Redes Neuronales convolucionales (CNN)

Shrestha y Mahmood [Shrestha y Mahmood, 2019] definen las Redes Neuronales como una técnica de Machine Learning inspirada en el sistema nervioso central de humanos, su estructura se caracteriza por tener unidades de procesamiento distribuidas en capas de entrada, capas ocultas y capas de salida; los nodos que tienen cada una de estas están interconectados de forma adyacente. Usualmente estas redes tienen aplicaciones muy diversas que abarcan desde el reconocimiento de patrones y clasificación hasta la visión por computadora, procesamiento de lenguaje natural, regresión, entre otros [Alarcón, 2025].

Entre las Redes Neuronales se encuentra un subgrupo denominado Redes Neuronales Convolucionales (CNN), su factor diferenciador radica en que identifica peculiaridades con estructuras de convolución. Su arquitectura está inspirada en la percepción visual, donde una neurona biológica se equipa con una artificial y los núcleos representan receptores que detectan varias características; así las funciones de activación imitan la transmisión selectiva de señales neuronales cuando se supera determinado umbral [Alarcón, 2025].

A nivel de funcionamiento, las decenas o cientos de capas que puede tener una CNN aprende a detectar diferentes características de una imagen, aplicando filtros a las imágenes de entrenamiento de manera que la salida resultante de convolucionar cada imagen se emplea como estrada para la siguiente capa; estas características que obtienen los filtros pueden ser el brillo, bordes o alguna mucho más complejas que definen el objeto de forma singular [Mathworks, 2022].

A diferencia de una red neuronal tradicional, una CNN tiene pesos y valores compartidos para todas las neuronas ocultad de una capa específica, lo que quiere decir que todas estas detectan las mismas características en diferentes regiones de una imagen. Esto hace que la red sea tolerante a la traducción de objetos de una imagen independientemente de donde aparezcan [Mathworks, 2022].

La clasificación en una red ocurre después de aprender características en muchas capas; allí la penúltima capa está completamente conectada y genera un vector de K dimensiones, donde K es el número de clases que se pueden predecir, y contiene las probabilidades para cada clase de una imagen que se está clasificando; así la capa final es una capa de clasificación que brinda la salida de clasificación final [Mathworks, 2022].

Actualmente estas redes neuronales han sido utilizadas en muchos campos por su capacidad para descubrir y aprender características clave de datos de imágenes y series temporales, por eso enfoques como el análisis de imágenes médicas, el procesamiento de audios, la detección

de objetos y la generación de datos sintéticos, han utilizado estas herramientas como primera línea de uso. Además de que usualmente estas tareas requieren utilizar grandes cantidades de datos complejos, uno de los principales desafíos que abordan las CNN.

3.2.3. Preprocesamiento de datos

El preprocesamiento es una etapa de vital importancia en las aplicaciones de aprendizaje automático, debido a que el rendimiento de estos modelos depende de la cantidad, calidad y diversidad de datos. A grandes rasgos el procesamiento se puede interpretar como una transformación aplicada a un vector de datos, allí se busca conservar la información esencial de los datos, pero eliminando al menos uno de los problemas asociados al vector, así al terminar el proceso se obtiene un grupo de datos más importante y valioso que en su versión inicial [Maharana et al, 2022].

A partir de esta información se dice que el tratamiento de datos se lleva a cabo cuando se presentan las siguientes situaciones: se requiere comprender las características y la naturaleza de los datos, se busca extraer información más necesaria y significativa o cuando se quiere solucionar problemas relacionados con los datos (ruido, elementos indeseados, resolución, tamaño, cantidad de datos, entre otras).

Según esto las técnicas de preprocesamiento de datos se pueden clasificar en tres grupos principalmente como se muestra en la tabla 1.

Tabla 1. Técnicas de preprocesamiento de datos

Data Pre-processing Techniques		
Data Transformation	**Information Gathering**	**Generating new Information**
- Data Ordering - Data Modelling - Filtering, Aggregation and Summarization - Indexing and Ordering - Enrichment and Imputation - Modelling, Typecasting, Formatting, and Renaming	- Data Sampling - Data elimination - Data selection - Data visualization	- Data engineering - Time series analysis - Data fusion - Data augmentation - Dimensional analysis - Constructive induction

Elaborado por: Maharana et al, 2022

No todas las técnicas deben ser implementadas en un proyecto que requiera preprocesamiento, pues esto depende de las características de los datos y las necesidades del proyecto, a pesar de esto, hay un esquema general de cuatro pasos de preprocesamiento:

1. Proceso de limpieza de datos: El objetivo de esta fase es detectar los datos incorrectos o ruidosos y corregirlos o eliminarlos del conjunto de datos [Maharana et al, 2022], las técnicas implementadas pueden varias según el volumen de datos, sin embargo, a grandes rasgos se llevan a cabo la eliminación de datos irrelevantes o duplicados, la corrección de errores estructurales y valores faltantes.

2. Manejo del ruido: allí se implementan los filtros necesarios para garantizar que la relación de información relevante con respecto a la información ruidosa sea significativa, garantizando la calidad de cada dato.

3. Transformación de datos: allí hay una serie de herramientas que pueden ser utilizadas según las necesidades de los datos, por ejemplo, la normalización se usa para llevar las características presentes a una misma escala; el aumento de datos es de gran utilidad para expandir el tamaño de la muestra ayudando a la generalización de estos; también se encuentra el balanceo que ayuda a nivelar el volumen de los datos para cada una de sus clases. Además, se pueden encontrar técnicas para los datos individualmente, donde se busca eliminar artefactos no deseados o mejorar la calidad de visualización.

3.2.4. Balanceo de clases

Cuando se abordan trabajos de clasificación o predicción de clase, el desbalanceo representa un problema de alto impacto, dado que estadísticamente la clase que tiene un mayor peso tiende a acumular un mayor número de clasificaciones, lo cual es aún más relevante a medida que el desbalanceo es mayor [García Abad J., 2021]. Para afrontar estos desafíos habitualmente se implementan técnicas de balanceo de datos o *resampling,* que modifican la distribución original de la muestra eliminando instancias de la clase mayoritaria (*undersampling*) o creando nuevas instancias de la clase minoritaria (*oversampling*) [García Abad J., 2021].

En el u*ndersampling*, las técnicas se enfocan en igualar las distribuciones desbalanceadas eliminando instancias de la clase mayoritaria, las soluciones difieren en función a cómo decide qué instancias eliminar o cuáles conservar; según esto se pueden encontrar dos enfoques principales: aleatoriedad o métodos heurísticos fundamentados en el método de los k-vecinos. Allí se resaltan los métodos heurísticos dado que se mitiga el riesgo de pérdida de información relevante, conservando la variabilidad más importante vinculada a los datos; entre estos métodos se resaltan *Condensed Nearest Neighbours* (CNN), *Tomek Links, Edited Nearest Neighbours* (ENN) y *NearMiss;* a pesar de tener el mismo objetivo, todas difieren en ciertos factores como su agresividad con los datos y su funcionamiento [García Abad J., 2021].

En el *oversampling*, las técnicas operan sobre los datos aumentando el tamaño de la muestra original, allí el objetivo es replicar instancias en la clase minoritaria a partir de los casos existentes o reales, esta técnica representa un riesgo al generar información que no está incluida realmente en los datos. De igual manera que para la técnica anterior, se encuentran el método aleatorio o métodos heurísticos sintéticos como *Synthetic Minority Over-Sampling Techique* (SMOTE) y *Adaptative Synthetic Sampling* (ADASYN), siendo SMOTE el más referenciado en la literatura por su fundamento en los k-vecinos, donde replica las instancias de la clase minoritaria tomando cada caso e introduciendo nuevos ejemplos sobre las líneas de unión de sus k-vecinos más cercanos [García Abad J., 2021].

En cuanto a las técnicas híbridas de resampling, básicamente se aplican los métodos mencionados anteriormente de manera combinada, de tal manera que se eliminan instancias redundantes de la clase mayoritaria y luego de replican algunas de la clase minoritaria, las técnicas más comunes son *SMOTE+Tomek links* y *SMOTE+ENN*, las cuales difieren entre sí por la intensidad de limpieza de la clase mayoritaria, donde ENN se caracteriza por ser mucho más agresiva [García Abad J., 2021].

4. METODOLOGÍA

En este capítulo se describen las técnicas y procedimientos utilizados para desarrollar el modelo automático de tamizaje de cáncer de piel. La metodología adoptada, estructurada bajo el enfoque CDIO (Concebir, Diseñar, Implementar y Operar), abarca desde la recopilación y preprocesamiento de los datos hasta la selección, entrenamiento y ajuste del modelo basado en redes neuronales convolucionales. Cada etapa fue desarrollada siguiendo prácticas recomendadas en la literatura científica, y las decisiones técnicas se evaluaron y optimizaron para maximizar el desempeño del sistema propuesto.

4.1. PLANTEAMIENTO METODOLÓGICO

El cáncer de piel representa una problemática creciente en salud pública, especialmente por la dificultad en su diagnóstico temprano y la alta mortalidad asociada a melanomas. En Colombia, esta situación se agrava por el acceso limitado a dermatólogos en zonas rurales y por la alta dependencia de la experticia clínica para realizar diagnósticos precisos. Este escenario plantea una necesidad urgente de herramientas de apoyo que mejoren el tamizaje y detección temprana de lesiones cutáneas sospechosas.

Ante esta necesidad, se concibió el desarrollo de un modelo automático basado en inteligencia artificial (IA), utilizando redes neuronales convolucionales (CNN), con el objetivo de detectar lesiones cutáneas malignas a partir de imágenes dermatoscópicas. El enfoque se basa en aplicar técnicas de aprendizaje profundo para identificar patrones visuales asociados a diferentes tipos de cáncer de piel, minimizando la subjetividad diagnóstica y reduciendo los tiempos de respuesta.

El modelo está orientado en la mejora del proceso de tamizaje en contextos de atención primaria, facilitando el análisis preliminar de imágenes capturadas mediante dispositivos dermatoscópicos. Su evaluación en imágenes correspondientes a la población colombiana busca validar su aplicabilidad en el contexto local, teniendo en cuenta los fototipos de piel predominantes en el país, así como características clínicas propias. Esta fase inicial del proyecto se centró en la selección del enfoque computacional más adecuado para abordar la clasificación multiclase de lesiones cutáneas en imágenes dermatoscópicas; considerando criterios como la precisión diagnóstica, la capacidad de generalización del modelo y su viabilidad para una futura implementación en entornos clínicos reales.

4.2. DISEÑO DEL MODELO

En esta etapa se estructuró el modelo automático desde el diseño del dataset hasta la arquitectura de red basado en los requerimientos establecidos en la Tabla 2. Se seleccionaron técnicas específicas de preprocesamiento, balanceo de clases y evaluación comparativa de

diferentes arquitecturas de redes neuronales convolucionales, siguiendo prácticas recomendadas en la literatura. Esta fase permitió construir una base técnica sólida antes del entrenamiento.

Tabla 2. Requerimientos generales del modelo

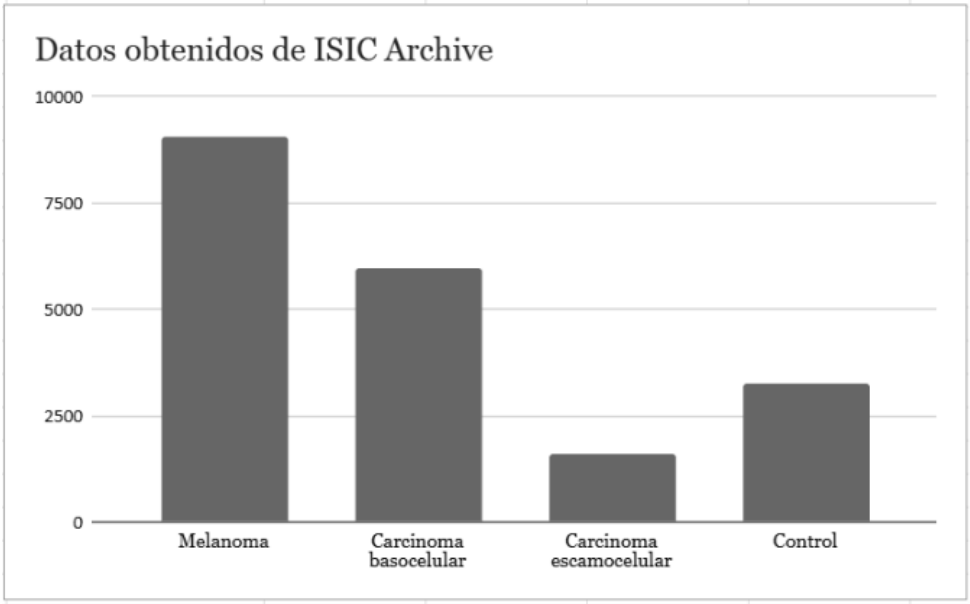
Requerimientos		
Funcionales	**Calidad**	**Restricción**
El modelo debe ser capaz de detectar y clasificar lesiones dermatológicas presentes en bases de datos públicas, mostrando la probabilidad asociada a las principales patologías cutáneas de interés. El modelo debe tener la capacidad de identificar tres tipos diferentes de cáncer de piel (melanoma, carcinoma basocelular y carcinoma escamocelular) a partir de imágenes dermatoscópicas. Se debe desarrollar una interfaz gráfica que integre el modelo, permitiendo al usuario cargar imágenes, ejecutar la predicción y visualizar los resultados de manera clara e intuitiva. El sistema debe procesar una imagen a la vez, asegurando la correcta asignación de probabilidades individuales y evitando conflictos en la ejecución.	El modelo debe alcanzar una precisión mínima del 85% con bases de datos internacionales para poder hacer las pruebas en pies colombianas. El valor de F1-Score debe mantenerse superior a 0,80, garantizando un equilibrio entre la precisión y el recall. El área bajo la curva ROC (AUC) debe ser mayor a 0.80, reflejando una buena capacidad de discriminación entre clases. La precisión del modelo no debe ser inferior al 80%, asegurando la correcta clasificación de los casos positivos detectados. El recall debe ser al menos del 80%, para minimizar la omisión de lesiones malignas.	La herramienta no tiene fines diagnósticos, sino que actúa como un sistema de apoyo al tamizaje y la orientación clínica. Las imágenes que analiza el modelo deben estar en formatos comunes como JPEG, PNG o TIFF. El modelo debe ser entrenado, validado y probado con imágenes dermatoscópicas. La evaluación del modelo debe incluir imágenes de bases de datos internacionales y muestras de diferentes fototipos de piel colombianos, con el propósito de comparar el desempeño y determinar su viabilidad local. Todas las imágenes utilizadas para entrenamiento y validación deben contar con lineamientos éticos y aprobación para su uso con fines académicos.

Elaboración propia

4.2.1. Recopilación y descripción de los datos

Los datos utilizados en este proyecto se obtuvieron de ISIC Archive (*International Skin Imaging Collaboration*) una asociación académica e industrial diseñada para utilizar imágenes digitales de la piel para ayudar a reducir la mortalidad por cáncer de piel [ISIC, 2024]. De allí se obtuvo un conjunto significativo de imágenes dermatoscópicas de contacto no polarizado para las clases Malignas: Melanoma, Carcinoma basocelular y Carcinoma de células escamosas; Benignas: Nevus, Dermatofibroma, Queratosis seborreica, Angiofibroma, Lentigo, entre otros. La cantidad de datos obtenidos se representan en la figura 1:

Figura 1. Datos obtenidos de ISIC Archive [Elaboración propia]



Cada imagen está asociada a un dataset donde se especifica: Isic ID, atribución, licencia copyright, edad aproximada, sexo, ubicación anatómica general, ubicación anatómica específica, clasificación entre benigno y maligno, biopsia concomitante, tipo de dermatoscopia, diagnóstico, diagnóstico 1, diagnóstico 2, diagnóstico 3, diagnóstico 4, diagnóstico 5 y tipo de confirmación del diagnóstico.

4.2.2. Preprocesamiento de las imágenes

El preprocesamiento de los datos recolectados es el primer paso y uno de los puntos más influyentes en el rendimiento de la generalización de un algoritmo de aprendizaje automático, allí se busca transformar los datos de manera que el modelo pueda revisar y analizar rápidamente sus características más relevantes. Este proceso abarca variedad de técnicas que sirven para comprender y extraer información significativa de los datos o solucionar problemas asociados a estos como el ruido y el desbalance [Maharana et al, 2022].

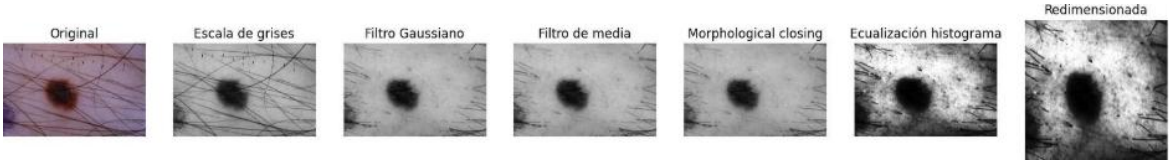
4.2.2.1. *Data Transformation*

Teniendo en cuenta lo mencionado, el preprocesamiento inició con lo que se llama '*Data Transformation*', allí se transforman los datos de manera que se resalten sus características más relevantes y se disminuya el ruido presente en cada una de estas, para esto se hizo una revisión de literatura de las principales técnicas de preprocesamiento para imágenes dermatoscópicas, en esta exploración se encontraron muchas combinaciones y tipos que arrojaron resultados funcionales a respectivos conjuntos de datos, por esta razón se implementaron las tres metodologías más comunes y efectivas de preprocesamiento y se evaluaron mediante el cálculo de la Relación Señal a Ruido (SNR) pues este término dentro del ámbito clínico y médico, se considera un concepto de suma importancia para comprender y optimizar el desempeño de diversas herramientas y técnicas de diagnóstico, a nivel general esta métrica relaciona la potencia de una señal, es decir la información relevante y la potencia del ruido de fondo (información irrelevante), esta herramienta funciona como un indicador de la calidad de la señal y en este caso permite evaluar si los métodos de preprocesamiento introducen ruido adicional, degradan la señal o la mejoran al resaltar bordes, texturas o estructuras clínicas relevantes [Clínica Universidad de Navarra, 2023]. En este contexto, esta métrica permite comprender cómo diferentes transformaciones afectan la capacidad diagnóstica de la imagen, dado que un SNR elevado se asocia con una mayor preservación de detalles útiles para el análisis dermatológico. A continuación, se describen las técnicas implementadas:

Técnica 1.

Esta primera metodología comenzó convirtiendo la imagen a escala de grises para eliminar la información de color irrelevante y centrarse en la textura y estructura de la lesión; luego, se aplicó una ecualización de histograma para mejorar el contraste, resaltando las diferencias entre las áreas claras y oscuras, lo que permite visualizar mejor los bordes de la lesión, posteriormente, se hizo un proceso de filtrado para suavizar la imagen y reducir el ruido sin perder detalles importantes para esto se utilizó un filtro de media 5x5 y un filtro Gaussiano 5x5. Para eliminar la interferencia de vellos en la piel, se aplicó la técnica de *morphological closing*, utilizando una estructura elíptica 15x15 que ayuda a cerrar pequeños espacios y borrar líneas delgadas [Vincent, 1994]. La Figura 2 muestra con una imagen el paso a paso de esta técnica.

Figura 2. Técnica 1 de Data Transformation en una imagen [Elaboración propia]



Técnica 2.

Esta técnica fue similar a la primera pues también busca resaltar la textura y la forma de la lesión por tal motivo, las imágenes se convierten una representación monocromática, a continuación, se aplicó un único filtro gaussiano con un kernel de 5x5 para suavizar la imagen y reducir el ruido de alta frecuencia; posteriormente se realizó una operación morfológica de cierre con un kernel de 40x40 por su utilidad para borrar elementos no deseados como vellos que circundan la lesión. Para optimizar el contraste y resaltar la lesión se utilizó CLAHE (*Contrast Limited Adaptive Histogram Equalization*) debido a su capacidad para ajustar localmente el contraste dividiendo la imagen en pequeñas regiones y ecualizando sus histogramas individualmente [Geek for Geeks, 2023]. Finalmente se empleó una técnica para mejorar la nitidez de la imagen, esta se llama '*Unsharp Masking*' una técnica que implica crear una imagen ligeramente desenfocada de un negativo original con un filtro gaussiano y mezclarla con la versión original [Instruments & Data Tools, 2021]. La figura 3 muestra con una imagen el paso a paso de esta técnica.

Figura 3. Técnica 2 de Data Transformation en una imagen [Elaboración propia]

Técnica 3.

A diferencia de las anteriores, esta técnica se basa en darle mayor importancia al color y la textura de la lesión, además de enfocarse en abarcar uno de los desafíos más relevantes en la clasificación de imágenes dermatológicas, los vellos, pues estas estructuras obstruyen la visibilidad de las características más importantes de la piel. Para aminorar esta dificultad se implementó el algoritmo 'DullRazor' diseñado específicamente para suavizar las imágenes y eliminar los vellos [Lee et al, 1997], su funcionamiento comienza convirtiendo la imagen RGB a escala de grises, luego se aplicó una operación morfológica denominada *BlackHat* utilizada para resaltar estructuras oscuras sobre fondos claros en una imagen con un kernel de 50 x 50 para ser agresivo con la eliminación de vellos, enseguida se segmentó la imagen en dos regiones binarias: una que identifica áreas con vello (marcadas en blanco) y otra que identifica áreas sin vello (marcadas en negro) determinando el valor mínimo de intensidad para considerar un píxel como parte del vello; finalmente con esa máscara se eliminan los vellos al rellenar las áreas marcadas como estos con información de los píxeles circundantes [Alarcón, 2025]. Dado que la imagen resultante se da en BGR se pasó a RGB y se aplicó la técnica de '*Unsharp Masking*' para mejorar la nitidez de la lesión, como se mencionó anteriormente. La figura 4 muestra con una imagen el paso a paso de esta técnica.

Figura 4. Técnica 3 de Data Transformation en una imagen [Elaboración propia]

NOTA. Algoritmo DullRazor tomado de Javier Velásquez (2020)

4.2.2.2. Balanceo de clases en el grupo 'Entrenamiento'

Para el entrenamiento de una CNN es muy importante que las clases estén balanceadas con el fin de evitar el sesgo del modelo hacia alguna de ellas, para estas situaciones existen técnicas como el *resampling* que tienen como finalidad igualar la distribución de las clases eliminando instancias de la clase mayoritaria (*undersampling*) o replicando instancias en la clase minoritaria a partir de la información de los casos existentes o reales (*oversampling*) [García Abad J., 2021]. Sin embargo, en este dataset se tiene un grupo (Melanoma) con muchos datos, un grupo (Carcinoma escamo celular) con muy pocos y otros dos grupos con valores intermedios entre ellos; dadas estas características lo más viable es implementar una técnica híbrida '*ensemble resampling*' los algoritmos más implementados y funcionales según la bibliografía son *SMOTE+ENN* y *SMOTE+Tomek links*, donde la principal diferencia entre estos es en la intensidad de la limpieza sobre la clase mayoritaria. Estas combinan el método *SMOTE* (*Synthetic Minority Over-sampling Technique*), que genera nuevas muestras sintéticas de las clases minoritarias, con métodos de limpieza (*Edited Nearest Neighbors* o *Tomek Links*) que eliminan muestras mal clasificadas o redundantes de las clases mayoritarias; el principal objetivo es mejorar la representatividad de las clases menos frecuentes, como el carcinoma escamo celular, sin aumentar excesivamente el ruido en los datos.

Dada la importancia de esta etapa se utilizaron ambas técnicas en el conjunto de datos 'Entrenamiento'. La implementación se realizó mediante la librería *imbalanced-learn* (*imblearn*) de python, ajustando SMOTE con 5 vecinos y los métodos de limpieza con sus configuraciones predeterminadas, ambas se evaluaron mediante tres factores y se selecciona el conjunto de datos con los resultados más funcionales. Métricas de evaluación para la técnica de balanceo:

Índice de balanceo de clases (IB); $IB = \frac{Tamaño\ clase\ mioritaria}{Tamaño\ clase\ mayoritaria}$

Pérdida de datos (P%); P (%) $= \frac{\#Imágenes\ descartadas}{\#Imágenes\ originales} \times 100$

Calidad del dataset (SNR); $SNR = 10 \times \log10 \left(\frac{Potencia\ señal}{Potencia\ ruido}\right)$

4.2.3. Red Neuronal Convolucional (CNN)

En esta sección, teniendo en cuenta las características del conjunto de datos y la revisión bibliográfica, se seleccionaron las arquitecturas más funcionales y utilizadas en modelos de aprendizaje. Además, se definieron los hiperparámetros propios de estas para poder llevar a cabo el entrenamiento.

4.2.3.1. Selección de la arquitectura

La elección de las arquitecturas de redes neuronales convolucionales (CNN) para este proyecto se fundamentó en un proceso sistemático que combinó revisión bibliográfica con análisis comparativo mediante una matriz de decisión ponderada. En una primera etapa, se realizó una búsqueda en bases de datos científicas como *IEEE Xplore, Scopus* y *PubMed*, identificando modelos que han mostrado un alto desempeño en la clasificación de imágenes médicas, con énfasis en estudios que emplean dermatoscopia digital.

De esta revisión se identificaron ocho arquitecturas, seleccionadas estratégicamente por su historial de rendimiento en tareas de clasificación de imágenes y su aplicabilidad en contextos médicos, además de la disponibilidad de modelos pre-entrenados, estas fueron: AlexNet, VGG16, Inception V3, EfficientNetB0, MobileNet, ResNet50, DenseNet121, Xception. Para seleccionar las arquitecturas para la etapa de entrenamiento, cada una fue evaluada en función de cuatro criterios: eficiencia computacional (tiempo de entrenamiento e inferencia), precisión reportada, capacidad de generalización del modelo (robustez y menor sobreajuste) y facilidad de implementación (documentación asociada a pesos pre-entrenados). La matriz desarrollada se muestra en la Tabla 3 a continuación con valores asignados de manera subjetiva entre 1 y 5 pero basados en la revisión bibliográfica realizada.

Tabla 3. Matriz de decisión ponderada para la elección de arquitecturas para la CNN

Arquitectura	Precisión (0.30)	Generalización (0.35)	Eficiencia computacional (0.20)	Facilidad de implementación (0.15)	Ponderado
AlexNet	3	3	4	5	3.5
VGG16	4	4	3	5	3.95
Inception V3	5	4	4	4	4.3
EfficientNet	5	5	4	4	4.65
MobileNet	4	5	5	4	4.55
ResNet50	5	5	4	5	4.8
DenseNet121	4	4	2	3	3.3
Xception	4	3	3	4	3.45

De acuerdo con el análisis comparativo, seis de las ocho arquitecturas analizadas presentaron un puntaje ponderado superior a 3.5, motivo por el cual fueron seleccionadas para la fase de implementación del modelo. Estas fueron:

AlexNet fue una de las primeras redes profundas que marcó un antes y un después en el campo de la visión por computador. Introducida en 2012, su arquitectura consta de ocho capas entrenables: cinco convolucionales que se encargan de la extracción de características, seguidas por tres capas completamente conectadas. Utiliza la función de activación ReLU, lo cual acelera el proceso de convergencia, y emplea *max pooling* para reducir la dimensionalidad espacial entre algunas capas. Además, incorpora técnicas como d*ropout* en las capas densas para prevenir el sobreajuste. Esta arquitectura, aunque relativamente simple comparada con modelos actuales, ofrece una base sólida y es útil como referencia por su claridad estructural [Shanthi & Sabeenian, 2019].

VGG16 es una red más profunda y simétrica que AlexNet, reconocida por su arquitectura uniforme, está compuesta por 13 capas convolucionales organizadas en bloques, seguidas de tres capas densas. Todas las convoluciones son de tamaño 3x3, lo que permite mantener una granularidad fina en la extracción de características [Mascarenhas & Agarwal, 2021]. Después de cada bloque de convoluciones se aplica *max pooling* para reducir las dimensiones. Las capas densas al final de la red están compuestas por dos bloques de 4096 neuronas cada una, antes de llegar a la capa de salida. La consistencia de su diseño la hace funcional para tareas de transferencia de aprendizaje, facilitando la adaptación a nuevos dominios en la imagenología médica [Mascarenhas & Agarwal, 2021].

Inception V3 es una red más sofisticada, desarrollada con el objetivo de reducir el costo computacional sin sacrificar profundidad ni rendimiento. Su arquitectura está basada en bloques Inception, los cuales combinan en paralelo múltiples operaciones (convoluciones 1x1, 3x3, 5x5 y pooling), cuyas salidas se concatenan para capturar información a diferentes escalas. Para optimizar la eficiencia, la red implementa técnicas de factorización de convoluciones, como dividir una 5x5 en dos 3x3, y reduce la dimensionalidad mediante convoluciones 1x1. También incluye mecanismos auxiliares de aprendizaje y normalización por lotes que mejoran la estabilidad del entrenamiento [Sam et al, 2019].

EfficientNet, por su parte, representa una nueva generación de redes optimizadas mediante un método sistemático de escalado. Esta arquitectura introduce los bloques MBConv, basados en convoluciones invertidas y separables, que permiten conservar información contextual con bajo costo computacional, además, incorpora mecanismos de atención tipo *squeeze-and-excitation*, que permiten ajustar dinámicamente la importancia de los canales de características. Su diseño permite escalar la red en profundidad, ancho y resolución de manera balanceada, logrando así una alta precisión con una carga computacional significativamente reducida [Putri et al, 2025].

MobileNet fue diseñada específicamente para operar en dispositivos móviles, donde la eficiencia es crítica. Su arquitectura está basada en convoluciones separables por profundidad, lo cual divide la operación de convolución en dos etapas: una convolución por canal individual y otra que combina los canales, reduciendo así significativamente el número de parámetros. Incluye *batch normalization* y utiliza ReLU6 como función de activación,

optimizada para precisión en entornos de bajo consumo energético. Esta red es especialmente útil cuando se requiere desplegar modelos ligeros que mantengan una buena capacidad de clasificación [Putri et al, 2025].

ResNet50 es considerada por ciertos autores como revolucionaria en las redes profundas por la introducción de conexiones residuales. Esta arquitectura de 50 capas permite la propagación de información a través de atajos (*skip connections*), que evitan el problema de la desaparición del gradiente al permitir que capas profundas aprendan funciones de identidad. Aunque no se profundiza aquí en sus resultados, su diseño estructural lo hace ideal para tareas que requieren modelos robustos y muy profundos, como es el caso del análisis dermatoscópico [Mascarenhas & Agarwal, 2021].

Por otro lado, las redes no seleccionadas (DenseNet y Xception) son descritas a continuación con el fin de identificar las principales diferencias:

DenseNet121 es una arquitectura de red neuronal convolucional caracterizada por su conectividad densa entre capas, donde cada capa recibe como entrada las salidas de todas las anteriores dentro de un bloque. Este diseño promueve una mejor propagación del gradiente, una mayor reutilización de características y una reducción del número de parámetros en comparación con arquitecturas tradicionales. Gracias a estas propiedades, DenseNet logra un entrenamiento eficiente y una representación robusta de las características visuales, aunque puede aumentar la demanda computacional dependiendo del tamaño de los datos y la profundidad del modelo [Huang et al., 2017].

La red Xception por otro lado se basa en el principio de las convoluciones separables en profundidad, una extensión del diseño Inception que descompone las convoluciones estándar en dos operaciones: primero una convolución espacial por canal y luego una convolución punto a punto. Este enfoque permite una buena eficiencia en el uso de parámetros y mejora la capacidad del modelo para capturar relaciones complejas entre canales, logrando resultados competitivos en tareas de clasificación de imágenes. No obstante, su rendimiento depende del volumen y la calidad de los datos utilizados para el entrenamiento [Chollet, 2017].

4.2.3.2. Hiperparámetros de entrenamiento

Para culminar el diseño del modelo, se desarrolló una etapa de exploración de hiperparámetros con el fin de identificar la configuración que ofreciera un mejor rendimiento en términos de precisión, sensibilidad y capacidad de generalización. Esta fase es fundamental en el desarrollo de modelos de aprendizaje profundo, ya que pequeños cambios en los hiperparámetros pueden tener un impacto significativo en el comportamiento y los resultados de cada red [Belete & Huchaiah, 2022]. Para llevarla a cabo, se construyó una grilla de búsqueda (*grid search*) donde se evaluaron combinaciones específicas de dos tipos de parámetros: el tamaño del lote (batch size) y los hiperparámetros asociados a la técnica de data Augmentation.

El batch size determina cuántas imágenes se procesan simultáneamente antes de actualizar los pesos de la red en cada iteración del entrenamiento, es un parámetro crítico que influye tanto en el tiempo de convergencia como en la estabilidad del aprendizaje [Belete & Huchaiah, 2022]. Se probaron diferentes tamaños de lote para encontrar un equilibrio entre uso eficiente de la memoria y estabilidad del gradiente. Lotes más pequeños pueden introducir mayor ruido durante el aprendizaje, lo cual favorece la generalización, mientras que lotes grandes permiten una estimación más precisa del gradiente, pero pueden llevar a modelos menos robustos frente a variaciones en los datos [Belete & Huchaiah, 2022].

Paralelamente, se exploraron combinaciones de técnicas de aumento de datos (*data augmentation*) como una estrategia esencial para mejorar la capacidad del modelo de generalizar a nuevas imágenes [Mumuni & Mumuni, 2022]. Este procedimiento es particularmente relevante cuando se trabaja con imágenes médicas, dado que la variabilidad entre pacientes y condiciones de captura puede ser alta, de esta manera en la grilla se evaluaron cuatro transformaciones principales: rotación, zoom, flip horizontal y modificación del contraste. La rotación permite que el modelo reconozca lesiones en distintas orientaciones, replicando condiciones del mundo real donde las imágenes dermatoscópicas pueden no estar perfectamente alineadas. El zoom simula variaciones en la escala de las lesiones, promoviendo la robustez del modelo ante diferentes niveles de acercamiento durante la captura. La transformación de flip horizontal ayuda a evitar que el modelo aprenda patrones dependientes de la orientación lateral de las imágenes, mientras que el ajuste de contraste mejora la capacidad del modelo para detectar bordes y diferencias de textura, incluso en imágenes con baja calidad o iluminación desigual.

Cada combinación de estos parámetros fue evaluada de forma sistemática para todas las arquitecturas previamente seleccionadas, esta estrategia de validación cruzada por grilla permitió comparar el desempeño de los modelos bajo condiciones controladas, garantizando que la configuración final no fuera producto del azar, sino resultado de una búsqueda rigurosa y metodológica. El resultado de esta etapa fue la selección de los hiperparámetros que maximizaban el desempeño en las métricas de evaluación sobre el conjunto de validación, estableciendo así una base sólida para el entrenamiento final y la posterior fase de ajuste fino (*fine-tuning*).

Cada combinación se evaluó con una arquitectura CNN estándar y se entrenaron durante un número reducido de épocas (10) para identificar patrones de rendimiento sin altos costos computacionales. La métrica principal para comparación fue la exactitud (*accuracy*) en el conjunto de validación.

Los hiperparámetros a evaluar en la grilla fueron extraídos de la bibliografía (estado del arte) según su funcionalidad en modelos de detección por procesamiento de imágenes médicas, según esto los valores probados para cada parámetro fueron:

- Tamaño del lote (*batch size*): 32 y 64
- Flip horizontal: True, False
- Rotación, zoom: 0.1, 0.2 y 0.3
- Contraste: 0.1 y 0.2

Durante el diseño del modelo se establecieron otros hiperparámetros relevantes que influyen directamente en la dinámica de aprendizaje y en la estabilidad del proceso de optimización, entre los que se encuentran el optimizador, la tasa de aprendizaje, la función de pérdida, los mecanismos de regularización y técnicas de parada temprana (*early stopping*).

Comenzando con el optimizador, se seleccionó el algoritmo Adam (*Adaptive Moment Estimation*), que se ha consolidado como una de las opciones más robustas y eficientes en el entrenamiento de redes neuronales profundas. Esta combina las ventajas de dos algoritmos ampliamente utilizados: AdaGrad, que ajusta la tasa de aprendizaje en función de la frecuencia de los parámetros, y RMSProp, que utiliza promedios móviles para estabilizar la actualización de pesos [Zhang,2018]. Esto permite que Adam mantenga una tasa de aprendizaje adaptativa para cada parámetro individual, lo cual resulta útil en modelos con muchos parámetros, como las arquitecturas profundas utilizadas en este estudio. El funcionamiento interno de Adam se basa en la estimación de los momentos de primer orden (la media del gradiente) y segundo orden (la varianza del gradiente), los cuales son actualizados de manera exponencialmente ponderada en cada iteración [Zhang, 2018]. De esta forma, el optimizador no solo considera el gradiente actual, sino también el historial de los gradientes pasados, lo que le otorga mayor estabilidad y eficiencia en trayectorias de optimización complejas.

Respecto a la función de pérdida, se empleó la *Categorical Crossentropy*, adecuada para tareas de clasificación multiclase con codificación *one-hot* en las etiquetas. Esta función calcula la divergencia entre la distribución de probabilidad real (etiquetas) y la predicha por el modelo, penalizando más fuertemente las predicciones incorrectas con alta confianza [Alarcón, 2025], su uso es estándar en problemas de clasificación donde las clases son mutuamente excluyentes.

Para mejorar la generalización del modelo y prevenir el sobreajuste (*overfitting*), se incorporaron mecanismos de regularización como *dropout* y *early stopping*. El *dropout* se aplicó en la capa densa del modelo con una tasa de 0.5, lo que implica que, durante cada paso de entrenamiento, el 50% de las unidades de dicha capa se desactivaron aleatoriamente, esto fuerza al modelo a no depender excesivamente de características específicas, promoviendo redes más robustas y generalizables.

Por último, se utilizó la técnica de parada temprana (*early stopping*) como un método de control del entrenamiento. Esta técnica monitorea el desempeño del modelo en el conjunto de validación, y detiene automáticamente el entrenamiento si no se observa una mejora en la métrica objetivo (en este caso, la pérdida de validación) después de un número determinado

de épocas consecutivas. En este proyecto, se estableció una paciencia de 3 épocas, con lo que se evitó continuar el entrenamiento cuando el modelo se ajustaba a los datos de entrenamiento, garantizando un uso eficiente de los recursos computacionales y una mejor generalización.

4.3. ENTRENAMIENTO E IMPLEMENTACIÓN

Con el conjunto de datos ajustado, las arquitecturas seleccionadas y los hiperparámetros establecidos se procedió a la implementación del modelo mediante transferencia de aprendizaje y ajuste fino cuyo orden y funcionamiento lógico se evidencia en el Anexo A Diagrama de flujo del modelo. En esta sección, se detallan los procesos de entrenamiento de las redes seleccionadas y las métricas utilizadas para evaluar su rendimiento.

4.3.1. Entrenamiento inicial

Para el entrenamiento de los modelos seleccionados se utilizó el enfoque de aprendizaje por transferencia (*transfer learning*), una estrategia que permite reutilizar el conocimiento adquirido por modelos previamente entrenados en grandes bases de datos, como ImageNet, para adaptarlos a una tarea específica [Alarcón, 2025].

Para cada arquitectura seleccionada (AlexNet, VGG16, InceptionV3, EfficientNetB0, MobileNet, y ResNet50), se construyó un modelo compuesto por dos partes: una base convolucional pre entrenada, encargada de extraer características jerárquicas de las imágenes, y una cabeza clasificadora personalizada, diseñada para adaptarse a la tarea específica de clasificación en cuatro categorías (melanoma, carcinoma basocelular, carcinoma escamo celular y lesiones benignas).

En esta etapa inicial, la base convolucional se congeló completamente, o sea, sus pesos no se actualizaron durante el entrenamiento. La decisión pretendía evitar el sobreajuste, especialmente dada la escasa cantidad de imágenes en comparación con el número de parámetros de estas arquitecturas profundas. Al congelar las capas convolucionales, el entrenamiento se concentró en las nuevas capas densas añadidas al modelo, que consistían típicamente en una capa de pooling global, una capa densa intermedia con activación ReLU, una capa *Dropout* con tasa del 50% para regularización, y una capa final densa con activación *softmax* para la clasificación multiclase.

El entrenamiento se realizó empleando imágenes redimensionadas según el tamaño requerido de entrada para cada modelo, organizadas en lotes de tamaño 64, según los resultados de la grilla de hiperparámetros. Adicionalmente se incorporaron todos los hiperparámetros adicionales mencionados anteriormente (*Data augmentation*, optimizador Adam, tasa de aprendizaje y función de pérdida). Inicialmente se estableció un número de épocas de 30 pero durante todo el proceso, se implementó la técnica de *Early Stopping* con el fin de detener el

entrenamiento automáticamente cuando no se evidencien mejoras significativas en la pérdida de validación durante tres épocas consecutivas evitando el sobre entrenamiento. Como estrategia de validación para este entrenamiento inicial se utilizó *hold-out* (80 entrenamiento, 20 validación) dado el volumen de datos.

Cada modelo construido pasó por este entrenamiento inicial, su funcionamiento en esta etapa se evaluó con las métricas definidas para analizar todos los modelos. (Véase en el numeral 4.3.3).

4.3.2. Fine-tuning

Tras completar la etapa inicial de entrenamiento con redes convolucionales pre entrenadas, se llevó a cabo una segunda fase conocida como *fine-tuning* o ajuste fino, con el objetivo de mejorar el rendimiento del modelo en el dominio específico de trabajo (imágenes dermatoscópicas). En la etapa previa, las arquitecturas empleadas se utilizaron como extractores de características, manteniendo congeladas las capas convolucionales del modelo base para evitar alterar los pesos previamente aprendidos en grandes conjuntos de datos como ImageNet.

Durante la fase de *fine-tuning*, se procede a descongelar parcialmente la red base, lo que implica permitir que algunas de las capas previamente entrenadas puedan seguir aprendiendo. En términos simples, esto significa que los pesos (los valores internos que determinan cómo la red interpreta las imágenes) de ciertas capas se actualizan nuevamente durante el entrenamiento mediante un proceso conocido como retro propagación (*back propagation*). En este caso, se optó por descongelar únicamente las últimas capas convolucionales del modelo pre entrenado, partiendo de la hipótesis de que estas capas aprenden patrones más específicos del problema que se quiere resolver [Vrbančič & Podgorelec, 2020]. Por ello, permitir que se ajusten con imágenes reales de lesiones en la piel mejora la capacidad del modelo para reconocer características clínicas relevantes.

El proceso de entrenamiento se desarrolla de la siguiente manera:

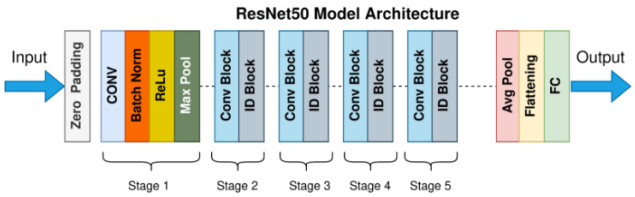
1. Las capas descongeladas inician con los pesos previamente aprendidos en el entrenamiento original.
2. Cada grupo de imágenes (lote) pasa por la red, donde se combinan los valores de las imágenes con los pesos de las neuronas en una etapa llamada propagación hacia adelante (*forward propagation*).
3. Los resultados llegan a las capas de activación (ReLU y softmax), que generan la predicción de la clase a la que pertenece cada imagen.
4. Una función de pérdida (*categorical crossentropy*) compara la predicción con la etiqueta real y calcula el error o pérdida obtenida en esa iteración.
5. Luego se calculan los gradientes, que indican en qué dirección y con qué magnitud deberían ajustarse los pesos para reducir el error, de acuerdo con la tasa de aprendizaje.

6. El optimizador ADAM utiliza esta información para modificar los pesos en las capas descongeladas, ajustándolos gradualmente durante el proceso de retro propagación.
7. Este ciclo se repite durante varias épocas de entrenamiento, hasta que los pesos se estabilizan en valores que permiten al modelo minimizar la pérdida y mejorar su precisión.

Dado que las tres arquitecturas seleccionadas para el entrenamiento final (ResNet50, MobileNet e InceptionV3) presentan estructuras internas significativamente distintas en cuanto a profundidad y distribución de capas, como se observa en las figuras 5, 6 y 7; fue necesario aplicar un proceso independiente de grid search para cada una. Se realizó una búsqueda de grilla para identificar cuántas capas debían descongelarse en cada modelo durante la fase de *fine-tuning*, ya que no era viable aplicar un mismo criterio generalizado. Esta variabilidad estructural exigió un cuidadoso análisis para cada caso, asegurando que el ajuste fino se realizara de forma eficiente sin comprometer la estabilidad del entrenamiento ni incurrir en sobreajuste.

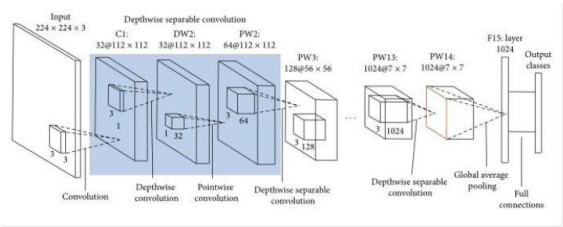
Para la ejecución de cada *grid search* se dividió el número total de capas en cuartiles y se tomó el valor promedio de cada cuartil para el análisis, de esta manera se puede evaluar rápidamente la tendencia central de la precisión dentro del conjunto de datos, además solo se llevarían a cabo siete épocas de entrenamiento para optimizar el gasto computacional y así tener un indicador de la tendencia en las métricas.

Figura 5. Arquitectura del modelo ResNet50 [Suvaditya Mukherjee, 2022]



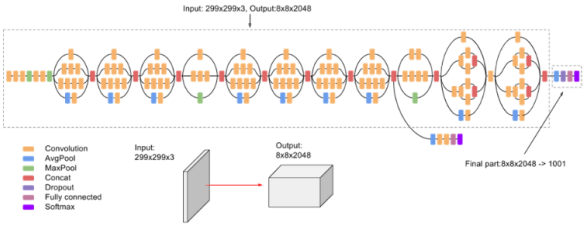
NOTA. Capas totales: 175

Figura 6. Arquitectura del modelo MobileNet [S. Amutha, 2024]



NOTA. Capas totales: 86

Figura 7. Arquitectura del modelo Inception V3 [Christian Szegedy, 2015]



NOTA. Capas totales:310

A partir de los resultados arrojados por el *grid search* se realizó un nuevo entrenamiento para cada modelo donde se descongelaron las capas correspondientes, con la finalidad de ajustar el modelo y maximizar los resultados. En esta fase se mantuvieron activas las estrategias de regularización como *dropout*, *early stopping* y *data augmentation*, para reducir el riesgo de sobreajuste dada la sensibilidad de esta etapa, así como el número de épocas (30).

Para la validación se implementó la estrategia: *hold-out* que permite medir el rendimiento del modelo en datos no vistos (capacidad de generalización a nuevos datos) y para evaluar el rendimiento real de un modelo de forma más robusta.

Finalmente se utilizan todos los resultados obtenidos de las métricas este proceso y se analizan para seleccionar el mejor modelo que será integrado con la interfaz.

4.3.3. Métricas de evaluación

Para la evaluación de los modelos durante el entrenamiento, se utilizaron las siguientes métricas:

- Matriz de confusión: tabla que desglosa las predicciones del modelo comparadas con las verdaderas clases.
- Exactitud: indica la proporción de predicciones correctas respecto al total de muestras.

  $Accuracy = \frac{VP + VN}{VP + VN + FP + FN}$; Sean VP Verdaderos positivos, VN Verdaderos negativos, FP Falsos positivos, FN Falsos negativos

- Precisión: Mide cuántas de las predicciones positivas realmente lo eran.

  $Pecision = \frac{VP}{VP + FP}$; Sean VP Verdaderos positivos, FP Falsos positivos

- Curvas ROC (*Receiver Operating Characteristic*) multiclase y AUC (*Area Under Curve*): permite analizar la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos de cada clase, además del área equivalente bajo cada curva asociada indicando la eficacia del modelo.

- *Recall:* mide cuántas de las verdaderas instancias positivas fueron correctamente clasificadas

$$Recall = \frac{VP}{VP + FN}$$

- *F1-Score*: es la media armónica entre precisión y *recall*, es un indicador del equilibrio del modelo predictor.

$$F1\ Score = 2 \times \frac{\mathrm{Pr}\,e\,cision\ \cdot\ Recall}{\mathrm{Pr}\,e\,cision\ +\ Recall}$$

- Curvas de entrenamiento y validación: muestran la evolución de una métrica (típicamente *accuracy o loss*) durante las épocas de entrenamiento y sirve para detectar el sobreajuste del modelo.

4.4. VALIDACIÓN EN ENTORNO REAL

La fase de operación contempla la integración del modelo seleccionado en una interfaz gráfica para su evaluación en imágenes de lesiones cutáneas recolectadas de población colombiana, incluyendo las pruebas funcionales y el análisis de los resultados.

4.4.1. Recopilación de imágenes de pieles colombianas

Las imágenes de pieles colombianas para validar el modelo desarrollado fueron tomadas de un artículo que tuvo como objetivo analizar la representatividad predictiva, analizando el sesgo racial en la detección del cáncer de piel mediante IA, donde hicieron una comparación de los dataset HAM10000 y BOSQUE Test, este último es un dataset recolectado de forma independiente por el grupo de investigación '*Infections and Clinical Dermatology Research Group*' de la Universidad el Bosque en junio del 2025 [Morales Forero, A. et al, 2025]. BOSQUE Test es un conjunto de datos de 110 imágenes dermatoscópicas de lesiones cutáneas pigmentadas de diferentes tipos, recopiladas en consultas dermatológicas y campañas de divulgación en Bogotá, Colombia. Según los metadatos de las imágenes el diagnóstico fue confirmado por histopatología. Todas las imágenes fueron capturadas con dermatoscopios de luz polarizada (DermLite+iPhone). El dataset tiene las siguientes características:

- **Sexo:** 97 femenino, 54 masculino
- **Grupos de edad:** de 20 a 90 años
- **Fototipos de piel de Fitzpatrick (Figura 8):**
    - Tipo II (justo, se quema fácilmente): 10 pacientes
    - Tipo III (marrón claro, quemaduras leves): 59 pacientes
    - Tipo IV (marrón moderado, rara vez se quema): 29 pacientes
    - Tipo V (marrón oscuro, muy raramente se quema): 7 pacientes
    - Tipo VI (pigmentación profunda, nunca se quema): 5 pacientes

Figura 8. Escala de fototipos de piel de Fitzpatrick [PMU Online Training, 2025]

Características de la lesión:

- **Naturaleza:** benigna o maligna (confirmada histopatológicamente)
- **Tamaño:** categorizado como ≤5 mm, 6–10 mm, 11–20 mm, >20 mm
- **Tiempo de evolución:** agrupado en categorías <1 año, 1 año, 2 años, 3–4 años, 5–9 años y 10 años o más
- **Sitio anatómico:** cabeza/cuello, tronco, extremidades o áreas acrales

Las clases comprendidas en este conjunto de datos son: Melanoma, Carcinoma de células basales, Queratosis actínica y carcinoma intraepitelial cutáneo (lesión precancerosa que puede evolucionar a carcinoma escamoso), Lesiones benignas tipo queratosis, Dermatofibroma, Nevus melanocítico y lesiones vasculares.

El dataset se encuentra público, anonimizado y depurado en el repositorio Harvard Dateverse: https://doi.org/10.7910/DVN/AQEPIN. El protocolo del estudio fue revisado y aprobado por el Comité de Ética Institucional de la Subred Integrada de Servicios de Salud Norte ESE y la Universidad El Bosque, esta base de datos cuenta con lineamientos éticos establecidos de conformidad con la normativa colombiana sobre protección de datos en salud y los estándares éticos internacionales, además de la aprobación e información a los pacientes para uso de los datos recolectados con fines académicos (de investigación y docencia) [Jaramillo Arboleda, A. et al, 2025].

4.4.2. Construcción de la interfaz gráfica

La construcción de la interfaz de usuario se desarrolló en dos etapas: diseño y ejecución. Ambos de vital importancia con el fin de garantizar la integración con el modelo desarrollado y su facilidad de uso para el usuario. Estos se describen a continuación:

4.4.2.1. Diseño conceptual de la interfaz

El diseño de la interfaz está centralizado en definir las bases estructurales y funcionales del sistema antes de pasar al desarrollo técnico. En este apartado se establecen los requerimientos que guían el diseño, se elaboran los bocetos iniciales que permiten visualizar la distribución

y jerarquía de los elementos, y se construyen diagramas de flujo que representan la lógica de navegación. Este enfoque metodológico asegura que la interfaz responda a criterios de usabilidad, simplicidad e intuición, al tiempo que traduce las necesidades del proyecto en un esquema.

Los requerimientos se establecieron con el objetivo de garantizar tanto la operatividad como la experiencia del usuario, se encuentra requerimientos funcionales y no funcionales respectivamente:

Requerimientos funcionales:

- El usuario debe poder subir una imagen en formato estándar en la interfaz.
- La interfaz debe mostrar una vista previa de la imagen cargada antes de procesarla.
- La interfaz debe enviar la imagen al modelo automático y devolver la categoría diagnóstica junto con las métricas asociadas a la predicción.
- Se debe permitir guardar los resultados en un archivo (ej. PDF, Excel o texto) para uso posterior.
- El tiempo de respuesta no puede ser superior a los 20 segundos.

Requerimientos no funcionales:

- Debe ser intuitiva, con íconos y textos claros.
- Colores neutros y tipografía clara.
- Las imágenes cargadas deben procesarse localmente sin quedar almacenadas en la aplicación manteniendo la privacidad del usuario.
- La interfaz debe poder ejecutarse en equipos de escritorio estándar.

Para la elección del nombre del modelo se buscó una propuesta con trasfondo que plasme el objetivo del proyecto pero que a la vez sea llamativo y sencillo, según esto, **NeoDerma** surge de la combinación de los términos “neo”, que significa nuevo o renovado, y “derma”, que hace referencia a la piel. De esta manera, el nombre refleja la esencia del proyecto: una propuesta innovadora que busca transformar la manera en que se realiza el tamizaje y diagnóstico temprano del cáncer de piel, integrando inteligencia artificial en beneficio de la salud pública. La gama de colores seleccionada se inspira en tonalidades asociadas a la confianza, la ciencia y la salud: los tonos azules transmiten seguridad y profesionalismo; el gris aporta neutralidad y equilibrio; mientras que el blanco refleja claridad, simplicidad y pureza.

Los siguientes bocetos se elaboraron con el fin de definir la distribución de los elementos y garantizar una experiencia de usuario sencilla e intuitiva, cumpliendo con los requerimientos establecidos anteriormente. Esto bocetos representan la disposición de cada pestaña de la interfaz: Inicio, Escanea, Acerca del Proyecto, Información Adicional y Términos y Condiciones. Cada uno muestra el encabezado (figura 9), el contenido central de cada pestaña plasmado en las figuras 11, 12,13, 14 y 15; además del pie de página (figura 10) y un diagrama de flujo de navegación plasmado en el Anexo B que representa la interacción entre las pestañas, asegurando una experiencia de usuario coherente y práctica.

Figura 9. Boceto encabezado de interfaz [Elaboración propia]

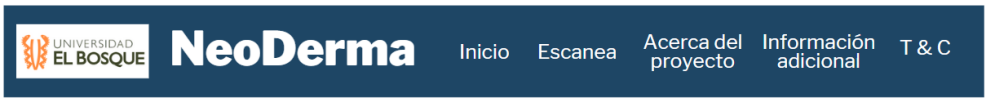


Figura 10. Boceto pie de página de interfaz [Elaboración propia]

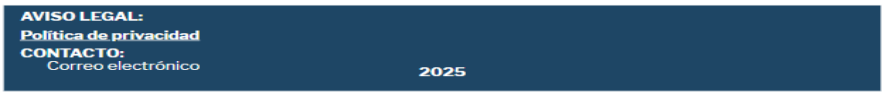


Figura 11. Boceto contenido de la pestaña 'INICIO' [Elaboración propia]

Figura 12. Boceto contenido de la pestaña 'ESCANEA' [Elaboración propia]

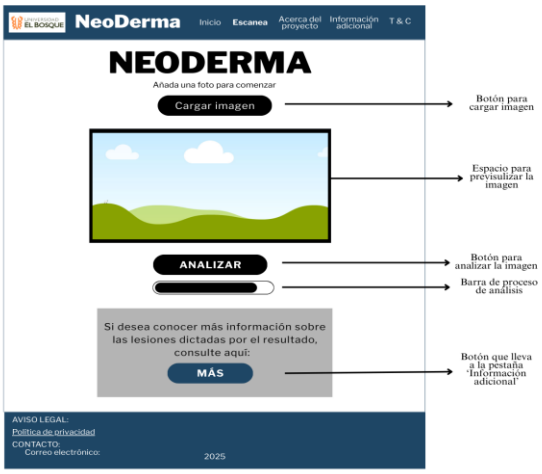


Figura 13. Boceto contenido de la Pestaña 'ACERCA DEL PROYECTO' [Elaboración propia]

Figura 14. Boceto contenido de la pestaña 'INFORMACIÓN ADICIONAL' [Elaboración propia]

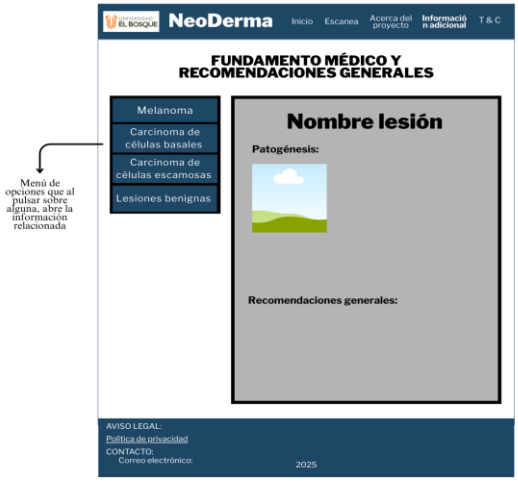
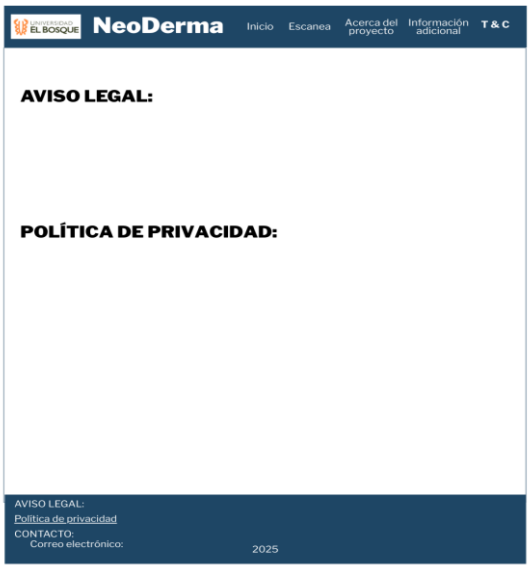


Figura 15. Boceto contenido de la pestaña 'TYC' [Elaboración propia]



4.4.2.2. Desarrollo de la interfaz

La construcción de la interfaz gráfica del modelo NeoDerma se desarrolló a partir de los bocetos y flujos de navegación previamente diseñados, con el objetivo de crear una herramienta intuitiva, clara y funcional para el tamizaje de cáncer de piel. Para el proceso de implementación se utilizó Visual Studio como entorno de desarrollo, gracias a su capacidad de integrar diferentes lenguajes y librerías en un mismo proyecto, su depurador avanzado y la facilidad que ofrece para gestionar módulos de manera organizada.

La interfaz se estructuró con HTML5, que proporcionó la arquitectura semántica de cada sección (Inicio, Escanea, Acerca del proyecto, Información adicional y Términos y Condiciones). Sobre esta base, se aplicó CSS(*Cascading Style Sheets*) que traduce Hojas de Estilo en Cascada, este lenguaje es utilizado para definir la presentación visual de los elementos HTML. Con este se dio forma al diseño visual; estableciendo tipografías, márgenes, proporciones y la paleta cromática definida (azul, gris y blanco). Para garantizar una visualización adecuada en distintos dispositivos, se incorporó el framework Bootstrap, que permitió implementar un diseño responsivo y aprovechar componentes prediseñados como menús de navegación, botones y tarjetas de información.

La parte dinámica de la aplicación se trabajó con JavaScript, lenguaje fundamental para dotar de interactividad a la interfaz. Su uso hizo posible programar botones y menús desplegables, validar formularios y, especialmente, gestionar la carga y visualización de imágenes dermatoscópicas en tiempo real. Al operar directamente en el navegador, JavaScript ofreció una integración fluida entre el contenido estructurado en HTML5 y los estilos de CSS, generando una experiencia de usuario clara y sin interrupciones.

Además, se integraron las librerías html2canvas y jsPDF, que añadieron una funcionalidad clave para la exportación de resultados. Html2canvas permitió capturar secciones específicas

de la interfaz y transformarlas en imágenes rasterizadas de alta resolución, mientras que jsPDF facilitó la incorporación de estas imágenes en documentos PDF portables. De este modo, el sistema no solo mostraba los resultados en pantalla, sino que también brindaba la opción de almacenarlos y compartirlos en un formato estandarizado y accesible, reforzando la utilidad clínica del prototipo.

La selección de estas herramientas no fue arbitraria: respondió a criterios de funcionalidad, compatibilidad y escalabilidad. HTML5, CSS y Bootstrap constituyen estándares consolidados para el desarrollo web, mientras que JavaScript aportó flexibilidad y control dinámico. Por su parte, html2canvas y jsPDF ofrecieron un valor diferencial al permitir la documentación inmediata de los resultados. En conjunto, estas tecnologías representaron un equilibrio entre robustez y facilidad de uso, asegurando que la interfaz cumpliera con los principios de usabilidad y pudiera adaptarse a las necesidades de entornos clínicos reales.

El desarrollo de la interfaz no se limitó a la integración de librerías y frameworks, sino que también contempló aspectos de arquitectura, diseño y usabilidad que resultan fundamentales en aplicaciones orientadas a la salud. En primer lugar, se definió una arquitectura cliente ligero, donde todas las operaciones se ejecutan directamente en el navegador del usuario. Esta decisión evita la necesidad de un servidor de almacenamiento, favoreciendo la protección de los datos clínicos, ya que las imágenes dermatoscópicas cargadas permanecen de forma local y no son transferidas a terceros.

Para conectar la interfaz gráfica con el modelo de clasificación NeoDerma, se implementó un backend desarrollado en Python mediante el microframework Flask, lo que permitió establecer una comunicación eficiente entre la aplicación web y el modelo de aprendizaje profundo. Flask actuó como intermediario entre el navegador y el motor de inferencia, gestionando las solicitudes HTTP generadas desde la interfaz (por ejemplo, la carga de una imagen dermatoscópica) y devolviendo las respuestas del modelo en tiempo real. El modelo, previamente entrenado y almacenado en formato .keras, fue cargado utilizando la librería TensorFlow/Keras de python, lo que posibilitó ejecutar el proceso de predicción directamente desde el servidor.

De esta manera, cada vez que un usuario cargue una imagen desde la interfaz, el sistema envía una solicitud al servidor Flask mediante el método POST, donde se ejecuta un script de preprocesamiento encargado de escalar, normaliza y adapta la imagen a las dimensiones esperadas por la red convolucional. Posteriormente, la imagen preprocesada se introducía en el modelo, que generaba las probabilidades de clasificación correspondientes a las cuatro categorías diagnósticas (carcinoma basocelular, carcinoma escamocelular, melanoma y control). Los resultados se estructuran en un objeto JSON y se devuelven al cliente, donde el código JavaScript los interpreta para actualizar dinámicamente la interfaz y mostrar el diagnóstico junto con su probabilidad asociada.

Esta arquitectura cliente-servidor ligera permite mantener un equilibrio entre velocidad de respuesta y modularidad, al separar la lógica de presentación (frontend) del procesamiento computacional (backend). Para optimizar el rendimiento, se definieron rutas específicas para la predicción y la descarga de resultados, lo que evita la sobrecarga del servidor y garantiza la estabilidad operativa del sistema.

4.4.3. Validación final del proyecto

Con el fin de validar la capacidad de generalización del modelo de clasificación en un contexto demográfico local, se diseñó esta fase de pruebas utilizando un conjunto de imágenes de pieles colombianas previamente recolectadas, según se menciona en el numeral 4.4.1. Esta etapa busca determinar el desempeño real del modelo al enfrentarse a variaciones en fototipos, tonalidades, y características morfológicas propias de la población nacional, constituyendo así una evaluación externa e independiente del entrenamiento principal. De esta manera evaluar qué tan viable es implementar este tipo de algoritmos en pieles colombianas e identificar posibles puntos de mejora en el desarrollo de modelos de aprendizaje con redes neuronales para el tamizaje de cáncer de piel, verificando si el modelo mantiene su rendimiento en un entorno clínico distinto al de su entrenamiento original, y determinando posibles sesgos o pérdidas de exactitud debidas a diferencias fenotípicas.

Para este proceso de validación se emplearon las 110 imágenes de pieles colombianas, la evaluación se llevó a cabo de manera retrospectiva y controlada, empleando el modelo final (Inception V3) ajustado mediante fine-tuning sobre la base internacional de referencia, en este caso ISIC.

Para esta fase final del proyecto se utilizó la versión del modelo con los mejores resultados en validación (menor pérdida y mayor accuracy), en este, cada imagen se procesa individualmente para generar una probabilidad de pertenencia a cada clase (CBC, CEC, Melanoma o Control), donde la clase predicha corresponderá a la categoría con la probabilidad más alta, estas predicciones se contrastan con las etiquetas diagnósticas previamente verificadas por criterio clínico o histopatológico.

Finalmente se calculan las métricas de desempeño, estimando los indicadores de exactitud global (accuracy), precisión, sensibilidad, F1-Score y una matriz normalizada por clase, permitiendo analizar la distribución de aciertos y errores entre las categorías evaluadas. Adicionalmente se definirán estas métricas por fototipo de piel permitiendo comprender el comportamiento del modelo a nivel local y realizar un análisis mucho más completo.

5. RESULTADOS Y DISCUSIÓN

En este capítulo se encuentran todos los resultados de cada una de las etapas planteadas en la metodología descrita anteriormente.

5.1. DISEÑO DEL MODELO

Según la metodología planteada en la fase de diseño, los resultados obtenidos se enmarcan en el preprocesamiento de las imágenes recolectadas para resaltar y ajustar sus características más relevantes, además de la exploración e hiperparámetros para el diseño de cada modelo a probar. A continuación, las secciones 5.1.1., 5.1.2. y 5.1.3. Muestran los resultados obtenidos en cada proceso.

5.1.1. Preprocesamiento: *Data transformation*

La métrica destinada a la evaluación de esta etapa y que además es el fundamento para la elección de la técnica más funcional en la transformación de los datos es la Relación Señal a Ruido (SNR) promedio de todo el conjunto de imágenes, calculada antes y después de aplicar cada una de las tres técnicas implementadas, de esta manera se espera evidenciar cuál técnica mejoró la calidad de todo el conjunto de imágenes. Los resultados obtenidos se muestran en la Tabla 4.

Tabla 4. Relación señal a ruido promedio en técnicas de preprocesamiento

Técnica de *Data transformation* implementada	**Relación señal a ruido**	
	SNR promedio de las imágenes originales	SNR promedio de las imágenes resultantes
Técnica 1	12.81 dB	6.89 dB
Técnica 2		15.46 dB
Técnica 3		21.68 dB

La SNR final obtenida para la técnica 1 fue de 6.89 dB, significativamente más baja que el valor inicial de 12.81 dB. Esto indica que, aunque la metodología aplicada resalta la textura de la lesión y elimina parcialmente los vellos, también introduce ruido en el proceso, posiblemente por la combinación del filtro de media con el filtro gaussiano y a la ecualización del histograma, pues esto causa cambios visuales muy drásticos a la imagen opacando la lesión; pese a la mejora en la visualización de bordes, la pérdida de SNR sugiere que esta técnica sacrifica calidad de imagen por mejora en el contraste y segmentación visual.

Por otro lado, la técnica 2 logró aumentar la SNR de 12.81 dB a 15.46 dB. La implementación de CLAHE para mejorar localmente el contraste, junto con la reducción de ruido mediante el filtro gaussiano y la nitidez añadida por la técnica de *'Unsharp Masking'*, muestra una mejora considerable en la calidad de la señal. El balance entre resaltar los detalles y mantener una buena relación señal a ruido hace que esta técnica sea prometedora para la detección automática de lesiones, sin embargo, la SNR promedio sólo aumentó un 20%, este es un indicador de que todavía las imágenes tienen cantidades de ruido significativas y podría ser necesario un preprocesamiento mucho más agresivo.

En el caso de la técnica 3 la SNR resultante alcanzó 21.68 dB, el valor más alto entre las tres técnicas evaluadas. El uso del algoritmo 'DullRazor' para eliminar vellos, combinado con la transformación de color y la nitidez final aplicada, permitió preservar y mejorar significativamente la calidad de la señal. Esto es un indicador de que la estrategia centrada en mantener la textura y el color de la lesión, mientras se eliminan obstrucciones visuales, ofrece una imagen más limpia y detallada, la convierte en la más adecuada para el contexto de este proyecto, lo cual se comprueba con el incremento de la SNR un 69%.

Al comparar los resultados de las tres metodologías, se observa que las técnicas que incorporan operaciones más específicas para la reducción de ruido y mejora de contraste logran resultados funcionales en términos de SNR; en el caso de la técnica 1 aunque fue efectiva para resaltar bordes, necesitaría ajustes adicionales para controlar la introducción de ruido adicional. La técnica 2 también muestra buenos resultados y podría ser una alternativa viable, sin embargo, la técnica 3 logró un equilibrio más adecuado entre mejora visual y preservación de la señal, por esta razón y la métrica que lo soporta, se seleccionaron las imágenes procesadas con esta técnica para las siguientes etapas del modelo de detección.

5.1.2. Preprocesamiento: Balanceo de clases

Los resultados de las métricas utilizadas para la evaluación de las dos técnicas implementadas para el balanceo de clases se muestran en la Tabla 5 a continuación.

Tabla 5.: Resultados de balanceo de clases con SMOTE+ENN y SMOTE+Tomek links

Métrica	**Técnica**	
	SMOTE + ENN	*SMOTE + Tomek links*
Índice de balanceo de clases (IB)	0.36	0.99
Pérdida de datos	51.22 %	+80.8 %
SNR promedio	21.75 dB	20.05 dB

NOTA: en SMOTE+Tomek links hubo ganancia más no pérdida de datos

A partir de los valores plasmados es posible identificar cuál técnica de balanceo es más funcional para este contexto. El índice de balanceo (IB) refleja cuán equitativa quedó la

distribución de las clases después del procesamiento al comparar los tamaños de la clase mayoritaria con respecto a la clase minoritaria; en esta métrica la técnica *SMOTE+Tomek Links* logró un IB de 0.99, lo que indica una distribución más balanceada que la obtenida con *SMOTE+ENN* (0.36).

En cuanto a la pérdida de datos, *SMOTE+ENN* mostró una pérdida del 51.22%, mientras que *SMOTE+Tomek Links* incluso presentó una ganancia de +80.8% en el número total de imágenes, sin embargo, esto sugiere que más de la mitad de los datos actuales son sintéticos. A pesar de esto, el promedio de la relación señal a ruido (SNR) fue ligeramente superior con *SMOTE+ENN* (21.75 dB) comparado con *SMOTE+ Tomek Links* (20.05 dB), esta pequeña diferencia sugiere que las imágenes resultantes de *SMOTE+ENN* conservaron mejor la calidad visual.

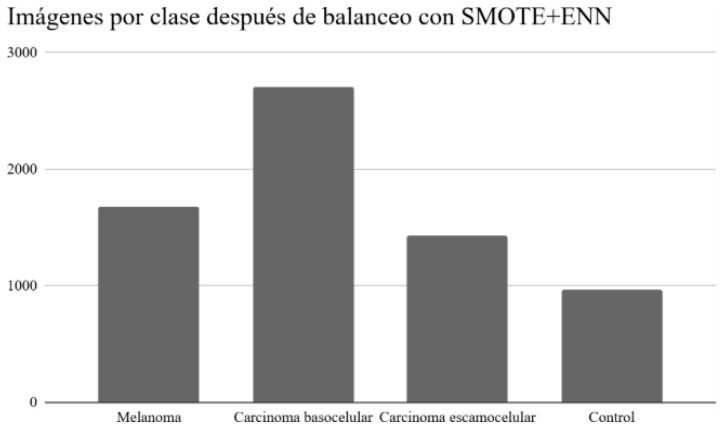
Uno de los aspectos clave para la elección de la técnica final fue la característica particular de *Edited Nearest Neighbors* (ENN), que se encarga de eliminar datos considerados irrelevantes o ruidosos [García Abad J., 2021]. Esto permite obtener un conjunto de datos más depurado, compuesto por las imágenes con las características más representativas de cada clase. Como resultado, el conjunto de datos no solo mejora en calidad, sino que también se reduce en tamaño, facilitando el procesamiento y disminuyendo los riesgos asociados a los límites computacionales, este factor es particularmente relevante debido al gran volumen de imágenes iniciales y la necesidad de optimizar el rendimiento del modelo.

Aunque *SMOTE+Tomek Links* logró un balance más uniforme, se priorizó la calidad de las imágenes y la reducción de complejidad computacional. Por ello, *SMOTE+ENN* fue la técnica seleccionada para continuar con la construcción del modelo.

En la Figura 16 se muestra la distribución final de las imágenes tras la implementación de SMOTE+ENN, allí se evidencia un aumento en la cantidad de imágenes para las clases minoritarias, especialmente carcinoma escamo celular y control, las más subrepresentadas. A pesar de esto, la clase carcinoma basocelular mantiene una mayor representación. Esto quiere decir que el balanceo con esta técnica proporcionó un conjunto de datos más limpio y manejable, conservando una calidad de imagen superior, esta decisión, aunque implicó sacrificar algo de balance entre clases, fue tomada con el objetivo de priorizar la precisión y eficiencia del modelo durante las siguientes etapas de entrenamiento y validación.

Cabe resaltar que esta distribución, aunque mejorada, deja entrever la posibilidad de seguir ajustando el balance con técnicas como '*Data Augmentation*' que permite nivelar aún más las clases realizando cambios de perspectiva a las imágenes para evitar el sobreajuste del modelo, lo cual sería lo óptimo teniendo en cuenta que las imágenes resultantes comprenden las características más representativas de cada clase.

Figura 16. Distribución de los datos con SMOTE+ENN [Elaboración propia]



NOTA: Melanoma 1679, Carcinoma basocelular 2708, Carcinoma escamo celular 1436, control 972

5.1.3. **Red Neuronal Convolucional: Grilla de hiperparámetros**

Las combinaciones evaluadas durante la grilla enfocada en analizar el efecto combinado del tamaño de batch (32 y 64) y parámetros de *data augmentation* aplicados al conjunto de entrenamiento, se muestran en la Tabla 6. Las transformaciones asociadas a *data augmentation* incluyeron volteo horizontal aleatorio (flip), rotaciones en el rango de 10° a 30°, aumentos de zoom (0.1–0.3) y ajustes de contraste (0.1–0.2). Dados los gastos computacionales implicados se decidió probar con siete combinaciones entre las posibles.

Tabla 6. Combinaciones de hiperparámetros evaluadas en la grilla

Trial	**Hiperparámetros**					**Resultado**
	Batch size	Flip	Rotation	Zoom	Contrast	Accuracy
1	32	Sí	10°	0.1	0.1	0.821
2	64	Sí	20°	0.2	0.2	0.839
3	32	No	30°	0.1	0.1	0.810
4	**64**	**No**	**10°**	**0.1**	**0.2**	**0.842**
5	32	Sí	20°	0.3	0.1	0.830
6	64	Sí	30°	0.2	0.2	0.838
7	32	No	10°	0.3	0.1	0.800

NOTA: la métrica de accuracy fue evaluada en validación hold-out 80-20

Entre las configuraciones probadas, la combinación con mejor rendimiento fue la número 4 que incluyó un *batch size* de 64, rotación de 10°, zoom de 0.1, contraste de 0.2 y sin

activación del flip horizontal, alcanzando una accuracy de validación de 0.842. Según estos resultados esta combinación fue seleccionada como punto de partida para el entrenamiento completo de los modelos.

5.2. IMPLEMENTACIÓN

En esta sección se muestran los resultados asociados a todos los entrenamientos realizados a los modelos, estos se muestran a continuación.

5.2.1. **Entrenamiento inicial**

La Tabla 7 presenta los resultados iniciales de entrenamiento de seis arquitecturas de redes convolucionales profundas, comparando métricas globales como accuracy, precisión, recall, F1-score y AUC (área bajo la curva ROC), las cuales métricas permiten evaluar el comportamiento general del modelo en la clasificación de imágenes dermatoscópicas, sin desglosar el rendimiento por clase. Teniendo en cuenta los hiperparámetros seleccionados se realizó el entrenamiento con los seis modelos, los resultados se muestran a continuación.

Tabla 7. Reporte inicial de entrenamiento para cada modelo

Arquitectura	No. Épocas	Accuracy	Precisión	Recall	F1 Score	AUC
VGG16	8	0.3068	0.3034	0.3068	0.2675	0.5151
ResNet50	16	0.8918	0.8932	0.8918	0.8925	0.9791
MobileNet	17	0.8374	0.8405	0.8374	0.8300	0.9679
EfficientNet	6	0.3974	0.1579	0.3974	0.1425	0.5824
InceptionV3	20	0.8175	0.8213	0.8175	0.8100	0.9604
AlexNet	11	0.3974	0.1579	0.3974	0.1425	0.4944

NOTA: valores considerados para el modelo general, no para cada clase.

A grandes rasgos se puede identificar como los modelos ResNet50, MobileNet e Inception V3 fueron aquellos con mejor desempeño global en todas las métricas, superando el 0.80 en cada una de ellas. Por otro lado, modelos como VGG 16, EfficientNet y AlexNet mostraron un rendimiento significativamente más bajo. Para interpretar estos resultados se debe considerar cuántas épocas lograron entrenar y las características de cada modelo.

El primer aspecto por destacar es la relación entre el número de épocas y el rendimiento de los modelos. Aquellas redes que obtuvieron mejores métricas, como ResNet50, MobileNet e InceptionV3, entrenaron durante un número moderado de épocas (16, 17 y 20 respectivamente), lo que sugiere una capacidad eficiente de aprendizaje sin necesidad de

recurrir a ciclos excesivos que podrían inducir sobreajuste. En contraste, los otros modelos alcanzaron un desempeño considerablemente inferior, incluso con menos épocas de entrenamiento, lo que se debe a una menor capacidad de generalización frente a las características particulares del conjunto de datos utilizado.

En cuanto a las métricas de desempeño, ResNet50 fue la arquitectura que logró los mejores resultados en todos los indicadores: una exactitud del 89.18%, una precisión del 89.32%, un recall del 89.18%, un F1-score de 89.25% y un AUC de 0.9791. Estos valores reflejan un modelo robusto y equilibrado, capaz de clasificar correctamente los casos positivos (alta sensibilidad), minimizar los falsos positivos (alta precisión) y mantener un desempeño general armónico (alto F1-score), además de una destacada capacidad de discriminación entre clases (elevado AUC). Este conjunto de resultados posiciona a ResNet50 como el modelo con mayor potencial para ser implementado en una herramienta de tamizaje automático en el contexto médico. Los modelos MobileNet e InceptionV3 también presentaron métricas consistentes, con valores superiores al 80% en precisión, sensibilidad y F1-score, y AUC cercanos a 0.96, lo que los convierte en alternativas viables, especialmente MobileNet si se considera la eficiencia computacional como criterio clave para una futura implementación en dispositivos portátiles o aplicaciones móviles. Estos modelos comparten características estructurales modernas y optimizadas para tareas de clasificación de imágenes, lo que explica su buen rendimiento en esta aplicación.

Las métricas elevadas que lograron alcanzar estos tres modelos son evidencia de su capacidad para ser utilizados como base para este proyecto, en particular porque su alta sensibilidad y ACU indican que estos modelos no solo detectan correctamente los casos malignos, sino que también lo hacen con alta seguridad, minimizando riesgos clínicos asociados; este tipo de comportamiento resulta muy funcional para herramientas de salud pública, donde fallar en la detección una lesión maligna podría tener consecuencias graves.

Por otro lado, los modelos VGG16, EfficientNet y AlexNet no lograron métricas satisfactorias. VGG16 presentó un rendimiento bajo en todas las métricas, con una exactitud apenas superior al 30%. EfficientNet y AlexNet obtuvieron resultados aún más bajos en precisión y F1-score, siendo de tan solo 15.79% y 14.25% respectivamente, lo que sugiere una alta tasa de errores en la clasificación, estos bajos desempeños podrían deberse a una falta de compatibilidad con los patrones visuales presentes en las imágenes dermatoscópicas o a la necesidad de realizar un ajuste más fino en sus hiperparámetros.

Dados los resultados obtenidos en esta fase del entrenamiento, el siguiente paso será llevar a cabo el *fine-tuning* con las tres mejores arquitecturas y enfocar la fase de validación con el ajuste final.

5.2.2. *Fine-tuning*

5.2.2.1. *Grid-Search*

Los resultados de *fine-tuning* se dividen en dos: los asociados a la búsqueda de grilla para identificar cuántas capas descongelar y los correspondientes al entrenamiento posterior con validación *hold-out*, estos se muestran a continuación.

En primer lugar, la tabla 8 refleja los resultados del *grid-search* para los tres modelos involucrados en el estudio:

Tabla 8. Valores evaluados en la grid search con sus resultados

Modelo	Número de capas descongeladas	Accuracy en validación
ResNet50	10	0.8168
	40	0.8513
	80	**0.8572**
	120	0.8462
MobileNet	20	0.5740
	40	0.7726
	60	**0.7741**
	80	0.6652
Inception V3	20	0.8416
	75	0.8682
	150	0. 8731
	225	**0.8821**

NOTA: Resultados con entrenamiento parcial a 7 épocas.

Los resultados de la búsqueda en grilla muestran comportamientos diferenciados según la arquitectura evaluada, por ejemplo, en el caso de ResNet50, se observa que el mejor desempeño en validación se obtuvo al descongelar 80 capas, alcanzando un accuracy de 0.8572. A medida que se incrementó progresivamente el número de capas descongeladas, la precisión tendió a aumentar de forma gradual, pero al descongelar el número total de capas la precisión disminuye, los valores plasmados reflejan que descongelando entre 40 y 80 capas acercándose más al tercer cuartil se obtendrán los mejores resultados, que se pueden optimizar incluso con más épocas de entrenamiento. Este comportamiento evidencia que la arquitectura de ResNet50 logra un mejor equilibrio cuando se ajusta una fracción significativa de sus capas profundas, lo que coincide con la naturaleza del transfer learning, donde gran parte de los pesos previamente entrenados en ImageNet ya representan características suficientemente generales para esta tarea.

El modelo MobileNet sigue con este tipo de comportamientos, alcanzando su precisión más alta (0.7741) al descongelar 60 capas, encontrando las métricas más altas entre el valor medio del segundo cuartil y el medio del tercero, con mejores resultados al estar más cerca al tercer cuartil. Este tipo de resultados son una guía muy útil para acercarse a los valores más funcionales de capas a descongelar para cada modelo, porque se pueden probar con diferentes valores que estén dentro de los rangos que mejor precisión arrojaron, permitiendo delimitar la búsqueda del número de capas a descongelar para obtener un resultado final donde los modelos alcancen su precisión más alta.

El mejor desempeño de InceptionV3 se alcanzó al descongelar 225 capas del modelo, lo cual permitió que el entrenamiento no se limitara únicamente a los clasificadores superiores, sino que involucrara también un ajuste fino de los bloques intermedios y profundos de la red. Este procedimiento favoreció la adaptación de los filtros convolucionales preentrenados en ImageNet hacia las características específicas de las imágenes dermatoscópicas, como variaciones en textura, coloración y bordes irregulares de las lesiones. En particular, la liberación de capas en el último cuartil evaluado posibilitó que los módulos Inception optimizaran la extracción de patrones multiescala relevantes.

Los resultados obtenidos evidencian diferencias en la capacidad de representación y en la velocidad de convergencia de cada modelo, atribuibles a sus características arquitectónicas y a la forma en que se ajustan a las particularidades del conjunto de datos dermatoscópicos. Por ejemplo, el modelo Inception V3 alcanzó la precisión más alta de los tres entrenados en tan solo 7 épocas de entrenamiento. Este comportamiento puede explicarse por la naturaleza de los módulos Inception, los cuales combinan convoluciones de diferentes tamaños en paralelo, permitiendo capturar simultáneamente patrones locales y globales en las imágenes. Esta propiedad de multiescala resulta especialmente útil en la clasificación de conjuntos [Qiang Z., He L. & Dai F.; 2019]. Además, la factorización de convoluciones grandes en operaciones más pequeñas contribuye a una convergencia más rápida y estable.

En contraste, ResNet50, aunque conceptualmente más profundo y con mayor capacidad teórica de representación gracias a los bloques residuales con “skip connections”, mostró un desempeño inferior en las primeras épocas, lo que sugiere un subentrenamiento bajo las condiciones iniciales aplicadas. Este modelo, al disponer de una mayor profundidad, requiere un ajuste más prolongado y la descongelación de un número mayor de capas para aprovechar plenamente su potencial.

Por su parte, MobileNet, diseñada para ser liviana y eficiente en dispositivos con recursos limitados, presentó la menor precisión, lo cual está asociado a su estructura basada en convoluciones separables en profundidad, que reducen significativamente la cantidad de parámetros a costa de limitar la capacidad de capturar patrones complejos. Aunque esta eficiencia es ventajosa en entornos móviles, en un conjunto de datos de mayor tamaño y con

características intrincadas, como el utilizado en esta investigación, puede resultar insuficiente para alcanzar desempeños comparables con InceptionV3 o ResNet50.

Estos resultados son un primer acercamiento al análisis del comportamiento de los modelos que se verá más adelante. En esta búsqueda InceptionV3 ofreció un balance más adecuado entre complejidad arquitectónica y capacidad de generalización, alcanzando altos niveles de precisión en menor número de épocas. No obstante, ResNet50 podría mejorar su rendimiento con un ajuste más exhaustivo de hiperparámetros y MobileNet se mantiene como una alternativa viable en escenarios donde la eficiencia computacional sea prioritaria sobre la precisión.

5.2.2.2. Entrenamiento con fine-tuning y elección del modelo

Conociendo los valores que pueden ser más favorables para cada modelo, se hizo un entrenamiento con *fine-tuning* para cada modelo y los resultados se muestran en las tablas 9 y 10. Adicional se muestra la curva de entrenamiento de cada uno en la figura 17.

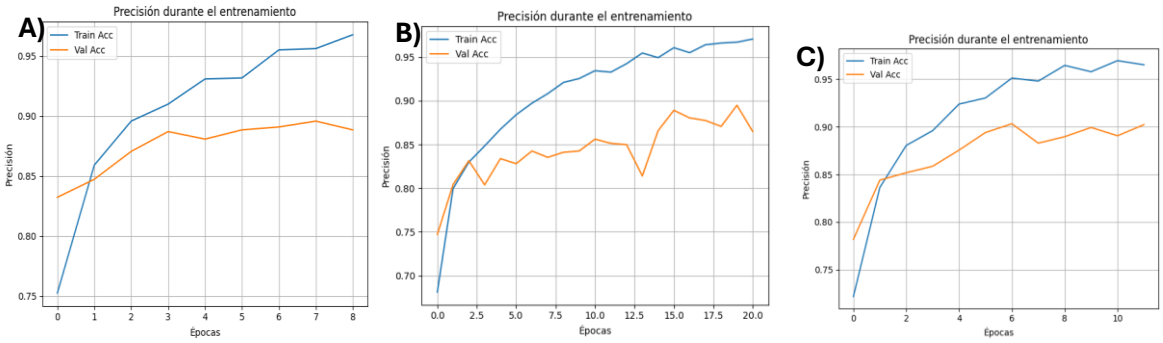
Tabla 9. Métricas de rendimiento global de los modelos con *Fine-tuning*

Métricas	**Modelos entrenados**		
	ResNet 50	MobileNet	Inception V3
Accuracy	0.8871	0.8891	0.9031
Precisión	0.8900	0.8892	0.9060
Recall	0.8871	0.8891	0.9031
AUC	0.9832	0.9841	0.9842

Tabla 10. Reporte de clasificación para cada clase

Clases	**Modelos entrenados**		
	ResNet 50	MobileNet	Inception V3
CBC	Precisión: 0.92 Recall: 0.87 F1-Score: 0.89	Precisión: 0.88 Recall: 0.92 F1-Score:0.90	Precisión: 0.94 Recall: 0.88 F1-Score: 0.91
CEC	Precisión: 0.80 Recall: 0.85 F1-Score: 0.82	Precisión: 0.88 Recall: 0.73 F1-Score: 0.80	Precisión: 0.84 Recall: 0.89 F1-Score: 0.86
Melanoma	Precisión: 0.94 Recall: 0.90 F1-Score: 0.92	Precisión: 0.93 Recall: 0.93 F1-Score: 0.93	Precisión: 0.94 Recall: 0.91 F1-Score: 0.93
Control	Precisión: 0.86 Recall: 0.93 F1-Score: 0.89	Precisión: 0.87 Recall: 0.94 F1-Score:0.91	Precisión: 0.86 Recall: 0.96 F1-Score:0.90

Figura 17. Curvas de entrenamiento de los modelos [Elaboración propia]



NOTA: Curvas de entrenamiento A) ResNet50, B) MobileNet, C) InceptionV3.

Para la evaluación del desempeño de los modelos se empleó un esquema de validación hold-out, dividiendo el conjunto de datos en un 80 % para entrenamiento y un 20 % para validación, con el propósito de medir la capacidad de generalización de cada arquitectura.

Los resultados globales de rendimiento (Tabla 9) evidencian un comportamiento sobresaliente de los tres modelos, con valores de *accuracy* superiores al 0.88 y áreas bajo la curva ROC (AUC) próximas a 0.98, lo que indica una alta capacidad discriminativa. Entre ellos, el modelo InceptionV3 alcanzó la mejor puntuación en todas las métricas: precisión = 0.9060, *accuracy* = 0.9031, *recall* = 0.9031 y AUC = 0.9842, consolidándose como la arquitectura con mayor poder predictivo. En comparación, ResNet50 y MobileNet obtuvieron valores muy similares entre sí (*accuracy* ≈ 0.887–0.889), aunque con ligeras diferencias en la estabilidad de las métricas.

El análisis por clase (Tabla 10) permite identificar con mayor detalle las fortalezas y debilidades de cada red. El modelo InceptionV3 presentó los mejores resultados globales con un *F1-Score* promedio de 0.90 y un comportamiento equilibrado entre precisión y *recall*. En este caso, la precisión refleja la capacidad del modelo para asignar correctamente una imagen a una categoría determinada, evitando los falsos positivos —es decir, diagnósticos erróneos de lesiones que en realidad pertenecen a otra clase—. Por su parte, el *recall* (o sensibilidad) mide la proporción de casos verdaderamente detectados dentro de todos los casos reales, expresando la habilidad del modelo para reducir los falsos negativos, aspecto crítico en la detección de patologías cutáneas donde omitir un caso puede tener consecuencias clínicas relevantes. El F1-Score, al ser la media armónica entre precisión y recall, sintetiza el equilibrio entre ambos factores, siendo particularmente útil cuando el conjunto de datos presenta desbalance entre clases, como ocurre en este estudio.

En la clase CBC, la arquitectura alcanzó una precisión de 0.94 y un *F1-Score* de 0.91, superando a las demás redes. Para la clase CEC, que históricamente representa mayor

dificultad debido a su variabilidad visual, InceptionV3 obtuvo un *F1-Score* de 0.86, evidenciando una mejor adaptación al conjunto de datos. En las clases Melanoma y Control, el modelo mantuvo un rendimiento homogéneo con valores de *F1-Score* entre 0.90 y 0.93, lo que refleja una buena generalización en todas las categorías.

En la clase CBC, InceptionV3 alcanzó una precisión de 0.94 y un F1-Score de 0.91, lo que indica un modelo altamente confiable en la identificación de este tipo de lesiones, con muy baja tasa de falsos positivos. Para la clase CEC, caracterizada por una mayor complejidad visual y heterogeneidad morfológica, el modelo logró un F1-Score de 0.86, superando a las demás arquitecturas y demostrando una mejor adaptación a la variabilidad del conjunto de datos. Este resultado sugiere que la arquitectura de Inception, con su capacidad para combinar convoluciones de diferentes tamaños, facilita la extracción de patrones multiescala, favoreciendo el reconocimiento de estructuras atípicas. Finalmente, en las clases Melanoma y Control, el modelo mantuvo un rendimiento homogéneo con F1-Scores entre 0.90 y 0.93, lo que refleja una adecuada generalización y una discriminación confiable tanto de lesiones malignas como de muestras normales. En conjunto, estos resultados evidencian que InceptionV3 no solo presenta un mejor desempeño cuantitativo, sino también una mayor estabilidad y robustez en la clasificación interclase, aspectos determinantes para su aplicación en entornos de apoyo diagnóstico automatizado.

Las curvas de entrenamiento y validación mostradas en la Figura 17 refuerzan estos hallazgos. En el modelo ResNet50, se observa una rápida convergencia del *accuracy* de entrenamiento que supera el 95 % hacia la octava época, mientras que la validación se estabiliza alrededor del 88 %, lo cual sugiere un ligero sobreajuste. MobileNet, en cambio, presenta oscilaciones más marcadas entre ambas curvas, indicativas de menor estabilidad y sensibilidad a los hiperparámetros, probablemente debido a su menor capacidad representacional. Finalmente, InceptionV3 muestra una evolución más uniforme, con una diferencia mínima entre entrenamiento y validación, alcanzando la mayor precisión en apenas seis épocas; esto evidencia una excelente capacidad de aprendizaje y generalización con un entrenamiento más eficiente.

Desde una perspectiva estructural, las diferencias de desempeño se explican por la naturaleza de cada arquitectura. InceptionV3 integra módulos de convolución paralelos con distintos tamaños de filtro [Abini, M. A.; Sathya Priya, S.; 2025], lo que permite captar información multiescala y combinar rasgos locales y globales de las lesiones cutáneas. Esta característica resulta particularmente ventajosa en imágenes dermatoscópicas, donde las irregularidades cromáticas y texturales son determinantes para el diagnóstico. Por su parte, ResNet50, basada en conexiones residuales, ofrece una gran profundidad y evita la degradación del gradiente, pero requiere un ajuste más fino del número de capas descongeladas y una mayor regularización para evitar sobre entrenamiento. MobileNet, diseñada para entornos de bajo

costo computacional, prioriza la eficiencia mediante convoluciones separables en profundidad, lo que reduce el número de parámetros, pero también limita su capacidad para modelar relaciones espaciales complejas.

En síntesis, los tres modelos demostraron un desempeño sólido en la clasificación de lesiones cutáneas; sin embargo, la superioridad de InceptionV3 se sustenta tanto en las métricas cuantitativas como en la estabilidad de su curva de validación. Su estructura modular, junto con el ajuste fino de 225 capas, posibilitó un aprendizaje más específico y una mejor adaptación a las características particulares de las imágenes analizadas. Estos resultados indican que InceptionV3 es la arquitectura más adecuada para la detección automática de lesiones cutáneas en población colombiana, ofreciendo un equilibrio entre precisión, sensibilidad y estabilidad de entrenamiento.

Los hallazgos obtenidos tienen implicaciones relevantes tanto en el ámbito metodológico como en el diagnóstico asistido por inteligencia artificial. En primer lugar, el rendimiento superior de InceptionV3 evidencia que los modelos con estructuras multiescala son más eficientes al abordar la variabilidad morfológica de las lesiones dérmicas. En segundo lugar, la aplicación del esquema de validación *hold-out* permitió estimar de manera confiable la capacidad de generalización, confirmando que el modelo no depende exclusivamente del conjunto de entrenamiento.

Desde una perspectiva práctica, la precisión alcanzada por InceptionV3 (≈ 90 %) lo posiciona como un modelo con potencial clínico para el apoyo al tamizaje temprano de cáncer de piel. No obstante, el desempeño ligeramente inferior en la clase CEC sugiere la necesidad de fortalecer esta categoría mediante estrategias adicionales de aumento de datos específicos o balanceo adaptativo, lo cual podría mejorar la sensibilidad sin comprometer la precisión.

5.3. INTERFAZ GRÁFICA

El desarrollo de la interfaz se realizó teniendo en cuenta los lineamientos establecidos en su diseño. Los resultados de este proceso se muestran a continuación.

La ventana ‘Inicio se muestra en la Figura 18, donde se puede evidencia la descripción general del modelo, sus objetivos y focos de interés, además está directamente vinculada a las pestañas ‘Escanea’ donde permite probar el funcionamiento del modelo y a la política de privacidad de la ventana ‘T &C’ (Figura 26) para conocer toda los términos legales que están asociados a este proyecto.

Figura 18. Resultado Ventana ‘INICIO’ [Elaboración propia]

En cuanto a la ventana correspondiente a la carga y análisis de las lesiones de piel que se desean probar en el modelo, se muestra en la Figura 19 que plasma como se logra visualizar la carga de una imagen específica con la barra de proceso del tiempo de análisis de

predicción. Una vez finaliza, automáticamente se descarga un archivo pdf con el informe del tamizaje que se evidencia en la Figura 20, con la probabilidad y la clase predicha.

Figura 19. Resultado Ventana 'ESCANEA' [Elaboración propia]

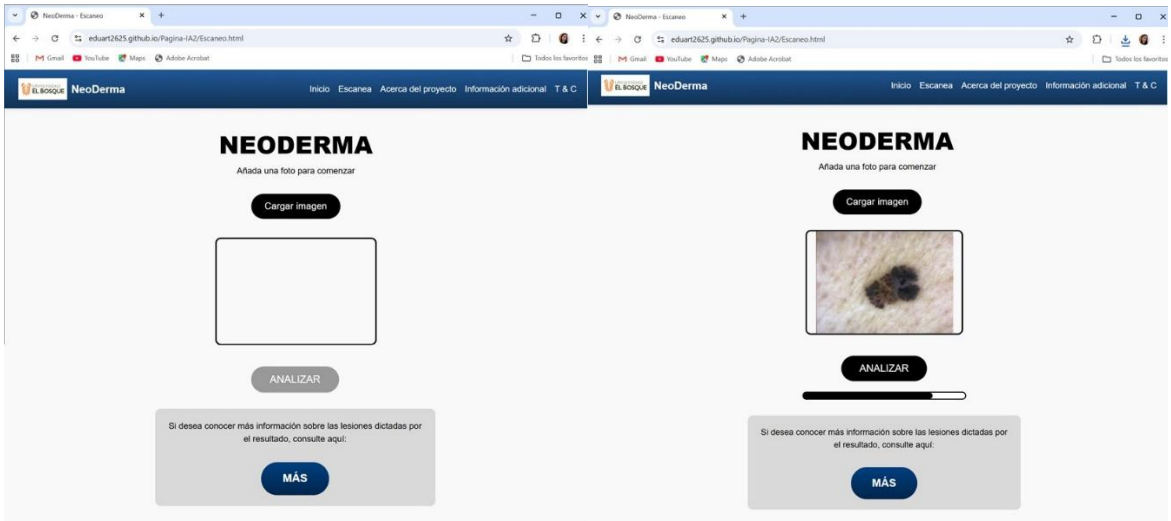
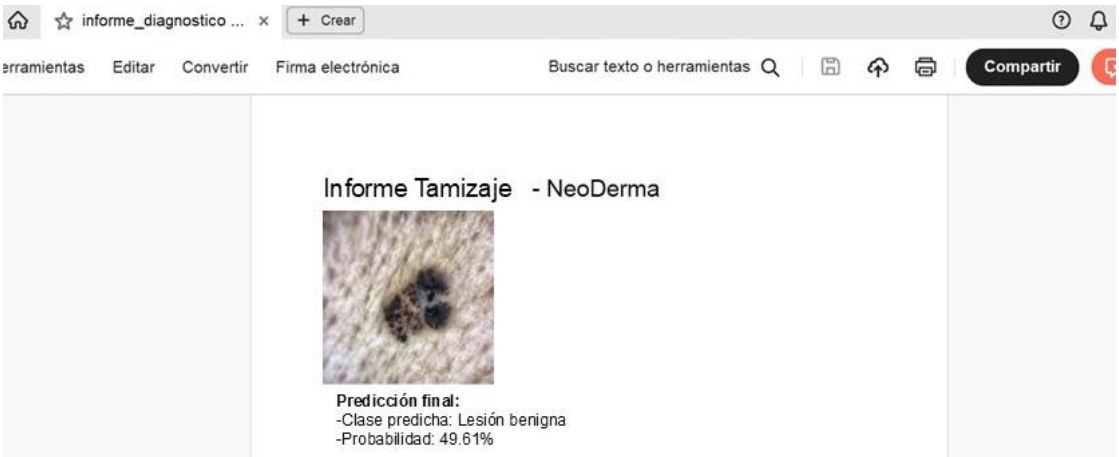


Figura 20. Resultado del informe de predicción del modelo con una imagen de prueba [Elaboración propia]



La pestaña 'Acerca del proyecto' se muestra en la Figura 21, corresponde a un abre bocas a este trabajo de grado mostrando de forma general el alcance, los resultados y las limitaciones asociadas, este contenido le permite al usuario comprender la relevancia de este proyecto y cómo abordar el problema identificado desde la bioingeniería.

Figura 21. Resultado Ventana ‘ACERCA DEL PROYECTO’ [Elaboración propia]

La pestaña ‘Información adicional’ comprende la patogénesis y unas recomendaciones generales so para cada una de las clases focales de este proyecto, esto se muestran en las Figuras 22, 23, 24 y 25.

Figura 22. Resultado Menú Melanoma [Elaboración propia]

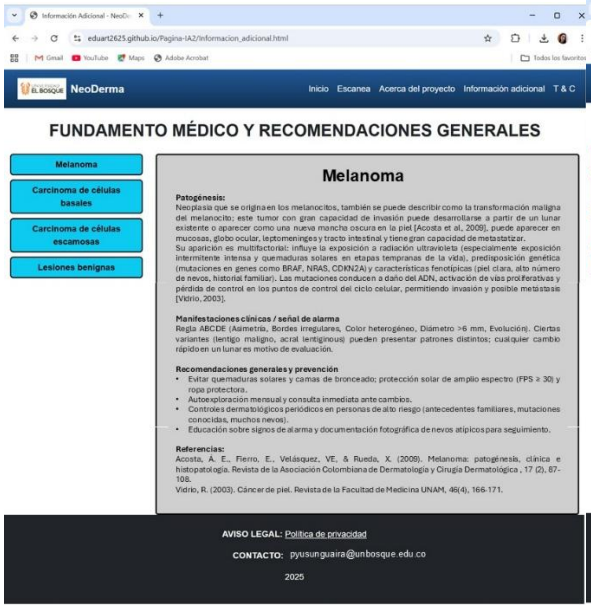


Figura 23. Resultado Menú CEC [Elaboración propia]

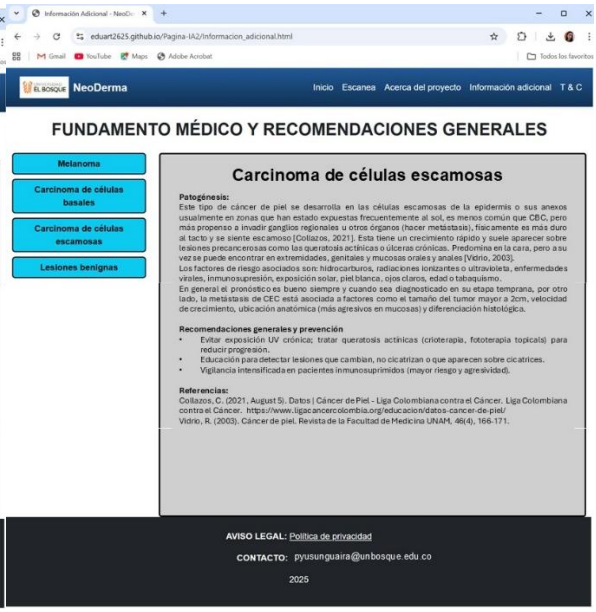


Figura 24. Resultado Menú CBC [Elaboración propia]

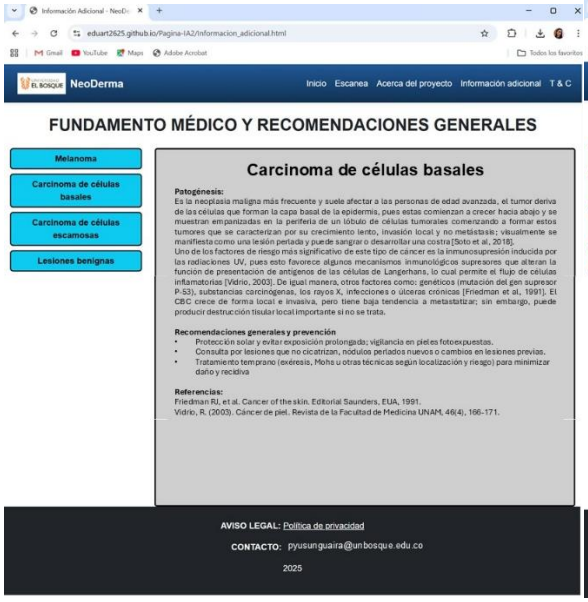
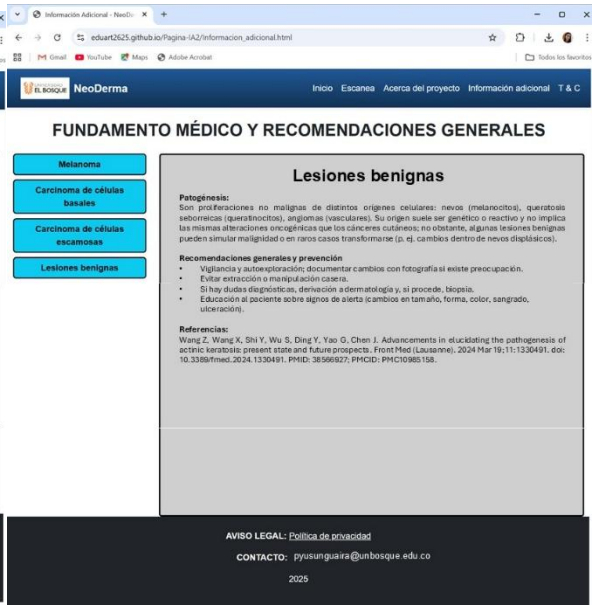
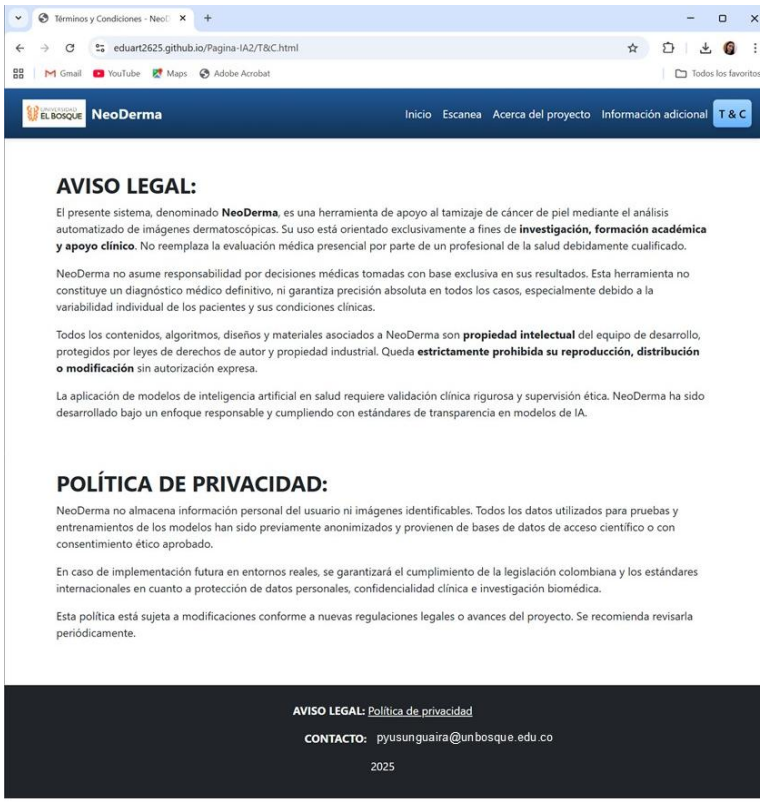


Figura 25. Resultado Menú Lesiones Benignas [Elaboración propia]



Por último, la Figura 26, muestra el resultado de la pestaña ‘ T & C’ donde se detalla el aviso legal y la política de privacidad para dar claridad al usuario sobre cualquier duda del tratamiento de datos y la responsabilidad de NeoDerma.

Figura 26. Resultado Ventana ‘T & C’ [Elaboración propia]



Para utilizar la interfaz correctamente se diseñó el manual del Anexo C Manual de usuario NeoDerma, que tiene como finalidad orientar al usuario en el uso adecuado de la herramienta y en la correcta interpretación de los resultados generados por el sistema. Aunque NeoDerma constituye un apoyo tecnológico valioso para el tamizaje temprano de lesiones cutáneas, es importante resaltar que sus predicciones se basan exclusivamente en el análisis automatizado de imágenes, por lo que existen limitaciones inherentes a la información disponible. En este sentido, se recomienda que, ante cualquier cambio en la lesión, dudas sobre la clasificación obtenida o signos clínicos de alarma, el usuario se dirija a una evaluación médica profesional, como lo específica el manual en sus limitaciones. Este manual acompaña al usuario en el proceso de uso responsable del sistema, promoviendo una comprensión adecuada de sus alcances y limitaciones.

5.4. VALIDACIÓN EN PIELES COLOMBIANAS

Esta fase de validación final en pieles colombianas comprende el análisis de la capacidad predictiva del modelo en variedad de fototipos de piel entre los que se encuentran los más representativos de la región.

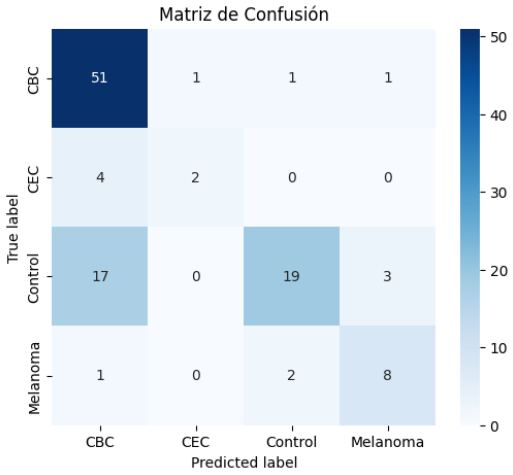
Según fuentes disponibles asociadas a estudios dermatológicos en Colombia, los fototipos de piel en escala de Fitzpatrick de tipo III, IV y V son los más frecuentes o predominantes en la población, pues en variedad de reportes de campañas poblacionales, encuestas o cohortes clínicas los sujetos de prueba corresponden en su mayoría a alguna de estas clasificaciones. A pesar de que no existe una estimación nacional representativa única por fototipo, es conveniente explicitar el enfoque de este estudio a los tipos de piel encontrados a partir de esta revisión bibliográfica [Flórez-Morales, I., et al., 2018; Restrepo Vélez, J. C., et al., 2001; Pardo-Zamudio, A. C., et al., 2020].

Los resultados asociados a este proceso de validación se muestran en la tabla 11 y en la figura 27 que muestran las métricas globales del modelo para ese conjunto específico:

Tabla 11. Métricas globales del modelo en validación con pieles colombianas

Clase	Precisión	Recall	F1-score
CBC	0.6986	0.9444	0.8031
CEC	0.6667	0.3333	0.4444
Melanoma	0.6667	0.7273	0.6957
Control	0.8636	0.4872	0.6230
Accuracy global		**0.7273**	

Figura 27. Matriz de confusión resultado global [Elaboración propia]



Según estos resultados, el modelo alcanzó una exactitud global de 0.7273, valor que refleja un desempeño moderadamente adecuado dentro del contexto de clasificación multicategórica. Sin embargo, las métricas por clase revelan un comportamiento heterogéneo que resulta crucial para la interpretación técnica y clínica de los resultados.

En el caso del carcinoma basocelular (CBC), el modelo obtuvo una sensibilidad elevada (recall = 0.9444), lo cual indica que logra identificar correctamente la mayoría de las lesiones de esta clase, aunque su precisión moderada (precisión = 0.6986) sugiere la presencia de un número apreciable de falsos positivos. En términos prácticos, esto implica que el modelo adopta un comportamiento conservador en la detección de CBC, privilegiando la minimización de falsos negativos, una estrategia que, aunque prudente desde una perspectiva clínica, puede aumentar la tasa de diagnósticos erróneos de esta categoría.

Por otro lado, para el carcinoma escamocelular (CEC) la sensibilidad fue sustancialmente baja (recall = 0.3333), lo cual significa que una gran proporción de estas lesiones no fueron reconocidas por el modelo, evidenciando una debilidad importante en su capacidad discriminativa hacia esta clase. Este comportamiento podría explicarse por la baja representatividad del CEC en la base de datos original, ya que esta categoría fue consistentemente la minoritaria. Si bien se aplicaron técnicas de balanceo de datos para compensar el desbalance de clases y mitigar posibles sesgos durante el entrenamiento, dicho procedimiento no garantiza una representatividad real de cada categoría ni asegura que las características intrínsecas de las lesiones de CEC hayan sido adecuadamente aprendidas por la red [Nguyen et al., 2022].

En el caso del melanoma, el desempeño fue intermedio y relativamente equilibrado, con un F1 de aproximadamente 0.70, mientras que para la clase control se observó un patrón opuesto al del CBC: alta precisión (0.8636) pero baja sensibilidad (0.4872), lo que indica que el modelo tiende a clasificar como “control” solo los casos en los que tiene mayor confianza, omitiendo un número significativo de verdaderos controles.

En conjunto, estas observaciones demuestran que la métrica global de exactitud enmascara diferencias sustanciales en el comportamiento del modelo entre las distintas clases y que, para fines clínicos, resulta más pertinente evaluar la sensibilidad y la precisión específicas que el promedio general. No obstante, como era previsible, el rendimiento obtenido en esta fase de validación difiere del alcanzado en etapas previas, ya que las imágenes empleadas presentan características particulares, especialmente en lo referente a la diversidad de fototipos de piel. En este sentido, resulta necesario analizar el desempeño del modelo de forma estratificada por fototipo, con el propósito de identificar posibles variaciones asociadas a las diferencias de pigmentación y representatividad de cada subgrupo. Y de esta manera comprender mejor la estabilidad y la representatividad predictiva del modelo en las diferentes tonalidades de piel. Los resultados de dicho análisis se presentan a continuación.

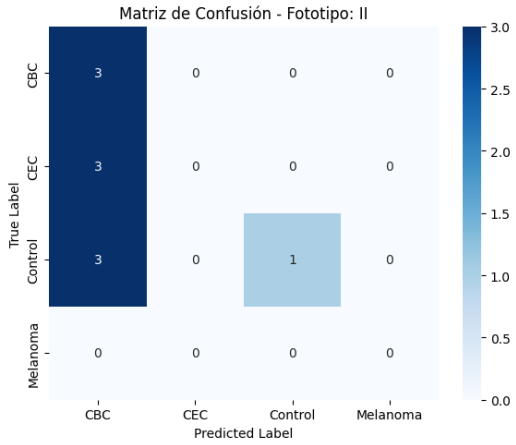
Tipo II

Los resultados de predicción del modelo para el fototipo de piel II se encuentran en la Tabla 12 y la Figura 28.

Tabla 12. Métrica local para fototipo de piel II

Clase	Precisión	Recall	F1-score	No. Muestras
CBC	0.33	1.00	0.50	3
CEC	0.00	0.00	0.00	3
Melanoma	0.00	0.00	0.00	0
Control	1.00	0.25	0.40	4
Accuracy		0.4000		

Figura 28. Matriz de confusión para fototipo de piel II [Elaboración propia]



En el fototipo II, correspondiente a pieles claras, los resultados fueron particularmente bajos y heterogéneos. En la clase CBC, el modelo alcanzó una precisión de 0.33 y una sensibilidad de 1.00, lo que significa que, aunque detectó todas las lesiones de esta clase, una gran parte de las predicciones positivas fueron erróneas. En contraste, para CEC y melanoma las métricas aparecen como 0; sin embargo, ello no implica una falla del modelo, sino la ausencia de muestras correspondientes a esas categorías dentro de este fototipo, lo que impidió el cálculo de dichas métricas. En la clase control, la precisión fue alta (1,00) y la sensibilidad baja (0,25), lo que refleja una tendencia conservadora en las predicciones. La limitada disponibilidad de ejemplos en este grupo genera una alta varianza en los resultados y restringe la posibilidad de evaluar con certeza el desempeño real del modelo en pieles de esta tonalidad.

Sin embargo, la baja eficacia observada en el fototipo II resulta particularmente interesante desde el punto de vista metodológico, ya que contradice parcialmente lo reportado en la literatura. En teoría, las pieles claras tienden a estar mejor representadas en la mayoría de los conjuntos de datos dermatológicos internacionales, como HAM10000 o ISIC, de modo que

el modelo debería exhibir mejor desempeño en estas condiciones. El hecho de que esto no ocurra en el conjunto colombiano sugiere la existencia de un fenómeno de cambio de dominio (*domain shift*) entre los datos de entrenamiento y los datos de validación [Chamarthi et al., 2023]. En otras palabras, aunque ambas poblaciones compartan el mismo fototipo nominal, las características visuales y contextuales de las imágenes —como los subtonos de la piel, las condiciones de iluminación, el balance de blancos o los equipos de captura— difieren de manera significativa. El fototipo II en Colombia puede presentar subtonos más cálidos o amarillentos, mientras que los fototipos claros europeos de las bases de datos originales suelen exhibir tonalidades más rosadas. A esta diferencia se suman las variaciones propias de la adquisición local. Tales factores alteran la distribución estadística de los valores de color y textura, afectando la capacidad del modelo para reconocer patrones aprendidos durante el entrenamiento. Este fenómeno ha sido ampliamente documentado en la literatura sobre *predictive representativity*, donde se demuestra que la representatividad demográfica no garantiza la representatividad predictiva, ya que la capacidad del modelo para generalizar depende de la proximidad entre las distribuciones de datos de entrenamiento y validación [Morales-Forero et al., 2025].

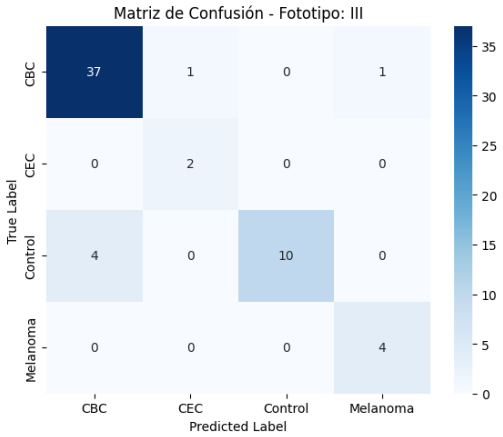
Tipo III

Los resultados de predicción del modelo para el fototipo de piel III se encuentran en la Tabla 13 y la Figura 29.

Tabla 13. Métrica local para fototipo de piel III

Clase	Precisión	Recall	F1-score	No. Muestras
CBC	0.90	0.95	0.92	39
CEC	0.66	1.00	0.80	2
Melanoma	0.80	1.00	0.88	4
Control	1.00	0.714	0.83	14
Accuracy		0.8983		

Figura 29. Matriz de confusión para fototipo de piel III [Elaboración propia]



El comportamiento del fototipo III, en cambio, representa el punto más sólido del modelo, con una exactitud de 0.8983 y valores F1 superiores a 0.80 en todas las clases, el modelo muestra un equilibrio adecuado entre precisión y sensibilidad, lo que indica que las características de las imágenes en este grupo son más compatibles con las representaciones internas aprendidas. Este desempeño sugiere que el conjunto de entrenamiento contenía una proporción importante de ejemplos visualmente similares a este fototipo, de modo que la red logró capturar con eficacia las texturas, contrastes y patrones cromáticos propios de este rango intermedio de pigmentación. La consistencia de los resultados en este grupo indica, además, que la arquitectura y los parámetros del modelo son adecuados para condiciones de luminancia media, lo cual explica por qué el rendimiento se degrada hacia los extremos claros y oscuros del espectro.

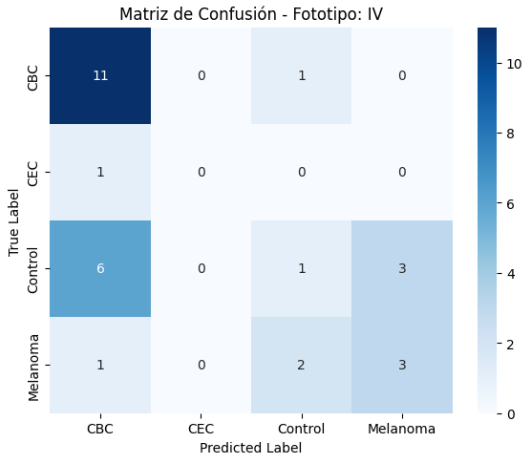
Tipo IV

Los resultados de predicción del modelo para el fototipo de piel IV se encuentran en la Tabla 14 y la Figura 30.

Tabla 14. Métrica local para fototipo de piel IV

Clase	Precisión	Recall	F1-score	No. Muestras
CBC	0.58	0.92	0.71	12
CEC	0.00	0.00	0.00	1
Melanoma	0.50	0.50	0.50	6
Control	0.25	0.10	0.14	10
Accuracy		0.5172		

Figura 30. Matriz de confusión para fototipo de piel IV [Elaboración propia]



El fototipo IV mostró resultados más heterogéneos. Aunque el modelo alcanzó una alta sensibilidad para CBC (0,92), las métricas para las demás clases fueron nulas o reducidas. Nuevamente, las métricas en cero no reflejan necesariamente errores del modelo, sino la

inexistencia de ejemplos válidos de ciertas clases en este grupo, lo que impide una evaluación completa.

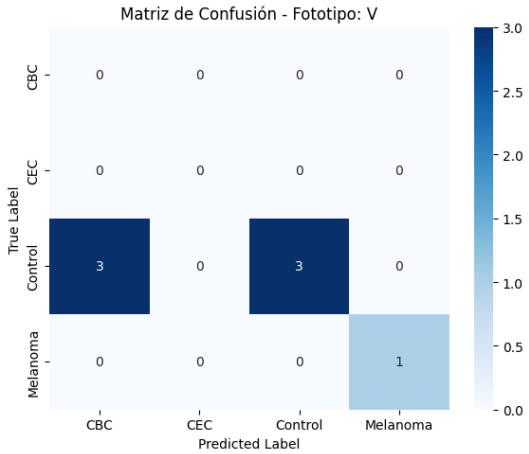
Tipo V

Los resultados de predicción del modelo para el fototipo de piel V se encuentran en la Tabla 15 y la Figura 31.

Tabla 15. Métrica local para fototipo de piel V

Clase	Precisión	Recall	F1-score	No. Muestras
CBC	0.00	0.00	0.00	0
CEC	0.00	0.00	0.00	0
Melanoma	1.00	1.00	1.00	1
Control	1.00	0.50	0.66	6
Accuracy		0.5714		

Figura 31. Matriz de confusión para fototipo de piel III [Elaboración propia]



El fototipo V presentó un comportamiento extremo, pues obtuvo métricas perfectas (precisión y recall = 1,00) en melanoma, pero valores nulos en CBC y CEC. Estos resultados, más que un rendimiento ideal o deficiente, son indicativos de la escasez de muestras por clase. Cuando la base de validación incluye una o pocas imágenes por categoría, un único acierto o error basta para producir métricas perfectas o nulas. En la clase control, el modelo alcanzó una precisión de 1,00 y una sensibilidad de 0,50, lo que mantiene el patrón observado en fototipos con soporte reducido.

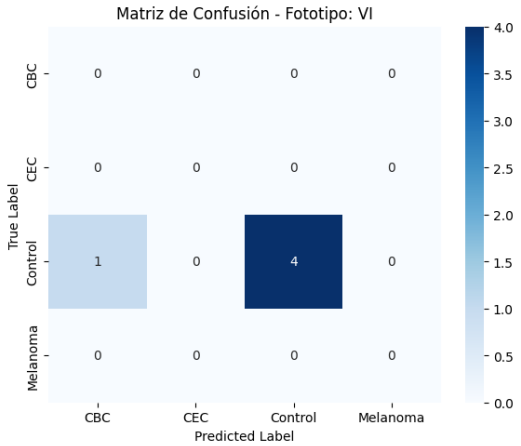
Tipo VI

Los resultados de predicción del modelo para el fototipo de piel VI se encuentran en la Tabla 16 y la Figura 32.

Tabla 16. Métrica local para fototipo de piel VI

Clase	Precisión	Recall	F1-score	No. Muestras
CBC	0.00	0.00	0.00	0
CEC	0.00	0.00	0.00	0
Melanoma	0.00	0.00	0.00	0
Control	1.00	0.80	0.89	5
Accuracy		0.80		

Figura 32. Matriz de confusión para fototipo de piel VI [Elaboración propia]



En el fototipo VI, únicamente la clase control mostró valores de desempeño (precisión = 1,00, recall = 0,80), mientras que las demás clases no pudieron ser evaluadas por falta de muestras representativas. Este escenario confirma que el conjunto de validación no fue balanceado en términos de distribución de clases por fototipo, lo cual limita el alcance inferencial del análisis.

Este patrón de métricas extremas, con valores perfectos o nulos, sugiere un problema de alta varianza estadística, muy probablemente derivado del bajo número de muestras disponibles para este fototipo en el conjunto de validación. Cuando el soporte por clase es pequeño, una única clasificación errónea puede modificar drásticamente los resultados, generando estimaciones poco confiables y que no reflejan el rendimiento real del modelo.

Cuando se analizan los resultados en conjunto, emerge una tendencia clara: el modelo exhibe un desempeño óptimo en el fototipo III, pero su rendimiento se torna errático y menos confiable en los demás grupos. Lo más llamativo es que esta degradación no sigue una progresión lineal asociada al tono de piel, sino que incluye también al fototipo II, que teóricamente debería mostrar mejores resultados por su predominio en los conjuntos de datos internacionales utilizados para el entrenamiento. Este hallazgo refuerza la idea de que la representatividad demográfica no necesariamente garantiza representatividad predictiva, pues la cantidad de ejemplos de un grupo no asegura que su distribución visual y contextual

esté adecuadamente representada. En otras palabras, el modelo puede haber sido expuesto a numerosas imágenes de piel clara, pero no necesariamente a ejemplos que reflejen la variabilidad de las pieles claras colombianas en términos de subtonos, iluminación y condiciones de captura.

Desde una perspectiva técnica, este fenómeno puede explicarse por un posible sobreajuste al rango medio de pigmentación, en el cual la red aprende representaciones discriminativas óptimas para niveles intermedios de contraste —como los propios del fototipo III— y pierde eficacia cuando el contraste es demasiado bajo o alto. Esta "ventana óptima de contraste" ha sido descrita en modelos de visión computacional que no implementan normalizaciones adecuadas de brillo y saturación [Buhrmester et al., 2021]. Asimismo, es posible que las imágenes del fototipo II en el conjunto colombiano hayan sido capturadas bajo condiciones de sobreexposición o variabilidad lumínica, ocasionando pérdida de detalle en la textura y los bordes de las lesiones. Estas diferencias, combinadas con los desbalances en el número de muestras por fototipo, reducen la capacidad de la red para identificar patrones clínicamente relevantes y generalizar de manera consistente.

En síntesis, los resultados de la validación confirman que el modelo mantiene una capacidad discriminativa satisfactoria en el rango de pigmentación intermedio, pero que su desempeño se deteriora en los extremos del espectro y, de forma inesperada, también en pieles claras locales. Este comportamiento demuestra que los modelos de aprendizaje profundo pueden presentar sesgos contextuales y distribucionales, más allá de los puramente fototípicos o raciales, lo que subraya la necesidad de construir conjuntos de datos locales con diversidad suficiente no solo en los fototipos cutáneos, sino también en las condiciones de adquisición, iluminación y presentación de las lesiones. Solo mediante un entrenamiento que refleje fielmente la heterogeneidad de la población objetivo será posible alcanzar una generalización robusta y equitativa, reduciendo el riesgo de sesgos clínicamente relevantes y garantizando una interpretación confiable del desempeño en todos los grupos poblacionales.

Así mismo durante esta fase final de validación los resultados demuestran la presencia potencial de sesgos inherentes en los sistemas de inteligencia artificial aplicados a la dermatoscopia, donde se presentan dos tipos de sesgo: uno originado por la falta de diversidad en los conjuntos de datos de entrenamiento, los cuales, en su mayoría, provienen de poblaciones con fototipos claros y características dermatológicas propias de regiones específicas como Europa o Norteamérica. Y otro dado por el tipo de dataset con el que se hizo la validación final en población colombiana, pues a pesar de que comprende todas las clasificaciones de fototipos de piel, este fue recolectado en una única área demográfica de Colombia (Bogotá) estableciendo una de las limitaciones más significativas de este proyecto.

Esta restricción geográfica afecta la capacidad de generalización del modelo frente a pacientes de otras regiones del territorio nacional, donde factores como la radiación solar, la

altitud, la ascendencia genética y las condiciones ambientales pueden influir en la apariencia de las lesiones cutáneas. En consecuencia, el modelo podría presentar variaciones en su desempeño diagnóstico al ser aplicado en contextos clínicos más diversos, lo que evidencia la necesidad de ampliar la base de datos con imágenes representativas de distintas zonas y grupos poblacionales para fortalecer su robustez y equidad en futuras implementaciones.

Finalmente, en comparación con los estudios revisados en el estado del arte, los resultados obtenidos en este proyecto reflejan avances consistentes, pero también evidencian limitaciones inherentes a los sistemas de apoyo diagnóstico basados en inteligencia artificial. Al igual que en investigaciones previas, donde la precisión del modelo dependió de la calidad y diversidad de los conjuntos de entrenamiento, este proyecto enfrentó desafíos asociados a la variabilidad de las imágenes y la homogeneidad demográfica, factores que han sido ampliamente reconocidos como fuentes de sesgo y error en otros desarrollos similares. Sin embargo, a diferencia de varios trabajos que emplean únicamente bases de datos internacionales con fototipos predominantemente claros y que sus métricas superan el 90% de exactitud, este estudio incorporó un proceso de validación con imágenes provenientes de población colombiana, lo cual permitió identificar de forma más precisa las limitaciones asociadas a la extrapolación del modelo, factor que se vio evidenciado en las métricas finales de validación. Asimismo, mientras que otros proyectos reportan restricciones vinculadas al tamaño de la muestra o a la ausencia de herramientas de interpretación clínica, este trabajo integró un sistema de tamizaje completo, pero aún enfrenta retos en la generalización hacia contextos rurales o con variabilidad étnica significativa. En conjunto, estas comparaciones demuestran que, si bien los resultados alcanzados son coherentes con la literatura, persisten desafíos técnicos y poblacionales que requieren ser abordados para fortalecer la utilidad clínica del sistema.

6. CONCLUSIONES

El estudio previamente expuesto enfocado en el análisis de la capacidad predictiva que tienen los modelos de aprendizaje para clasificar lesiones cutáneas malignas en la población colombiana, permitió demostrar el potencial de las redes neuronales convolucionales (CNN) en la clasificación de lesiones cutáneas a partir de imágenes dermatoscopias. Los resultados obtenidos durante la fase de implementación mostraron métricas globales superiores al 85% en precisión y exactitud, evidenciando un desempeño competitivo frente a estándares reportados en la literatura.

La fase de validación del modelo en pieles colombianas confirmó que la representatividad demográfica no garantiza por sí sola la representatividad predictiva, por lo que resulta esencial disponer de conjuntos de datos locales con suficiente diversidad en fototipos. Estos hallazgos refuerzan la necesidad de adaptar y validar los modelos de inteligencia artificial a las características propias de cada población antes de su implementación clínica.

El análisis estratificado por fototipo llevado a cabo en la etapa de validación reveló variaciones notables en el rendimiento del modelo, con una capacidad discriminativa adecuada en pieles de fototipo III y un deterioro significativo en los extremos del espectro de pigmentación (II y VI). Este comportamiento indica la presencia de sesgos asociados tanto a la distribución desigual de los datos como a las condiciones de adquisición de las imágenes.

El proyecto en general logra evidenciar el aporte de la bioingeniería al fortalecimiento del sistema de salud colombiano con el desarrollo de herramientas de diagnóstico complementarias, accesibles y no invasivas, que podrían contribuir significativamente a la detección temprana del cáncer de piel y a la reducción de las brechas en el acceso a la atención especializada. Sin embargo, estos deben ser entrenados con imágenes locales con el fin de evitar posibles sesgos asegurando la calidad de las predicciones.

7. RECOMENDACIONES

Para fortalecer y ampliar los resultados alcanzados, se proponen las siguientes líneas de acción y mejoras futuras:

Se recomienda dar continuidad a este proyecto mediante futuras investigaciones orientadas al perfeccionamiento del modelo y a la ampliación de las bases de datos utilizadas para su entrenamiento, incluyendo imágenes provenientes de diversas regiones y grupos poblacionales del país, esto permitiría mejorar la capacidad de generalización del sistema y reducir posibles sesgos en su desempeño. Asimismo, se aclara que el modelo requiere una segunda validación con imágenes de población colombiana, con el fin de fortalecer su precisión y confiabilidad en este contexto demográfico, teniendo en cuenta las consideraciones descritas en la discusión.

La relevancia de continuar con este trabajo radica en su potencial impacto social y sanitario, especialmente en la población del sector agropecuario, una de las más expuestas a la radiación solar y con mayores limitaciones para acceder a atención médica oportuna. En este sentido, el presente proyecto constituye un aporte valioso al desarrollo de soluciones tecnológicas aplicadas a la salud pública, promoviendo la equidad en el diagnóstico y el uso responsable de la inteligencia artificial en beneficio de las comunidades más vulnerables.

Para el perfeccionamiento del modelo y garantizar que pueda ser implementado en contextos reales de población colombiana, se pueden evaluar técnicas como reentrenamientos parciales (fine-tuning local) incorporando un subconjunto representativo de las imágenes colombianas, para mejorar la adaptación del sistema al dominio específico.

Validar el modelo con un conjunto de imágenes mucho más amplio que comprenda varias muestras de las diferentes lesiones para cada uno de los fototipos de piel. A pesar de que la muestra actual permitió tener un estimado de la capacidad predictiva del modelo por fototipo un análisis mucho más exhaustivo puede dar paso a reconocer puntos de mejora específicos del modelo como una clase específica en un fototipo dado.

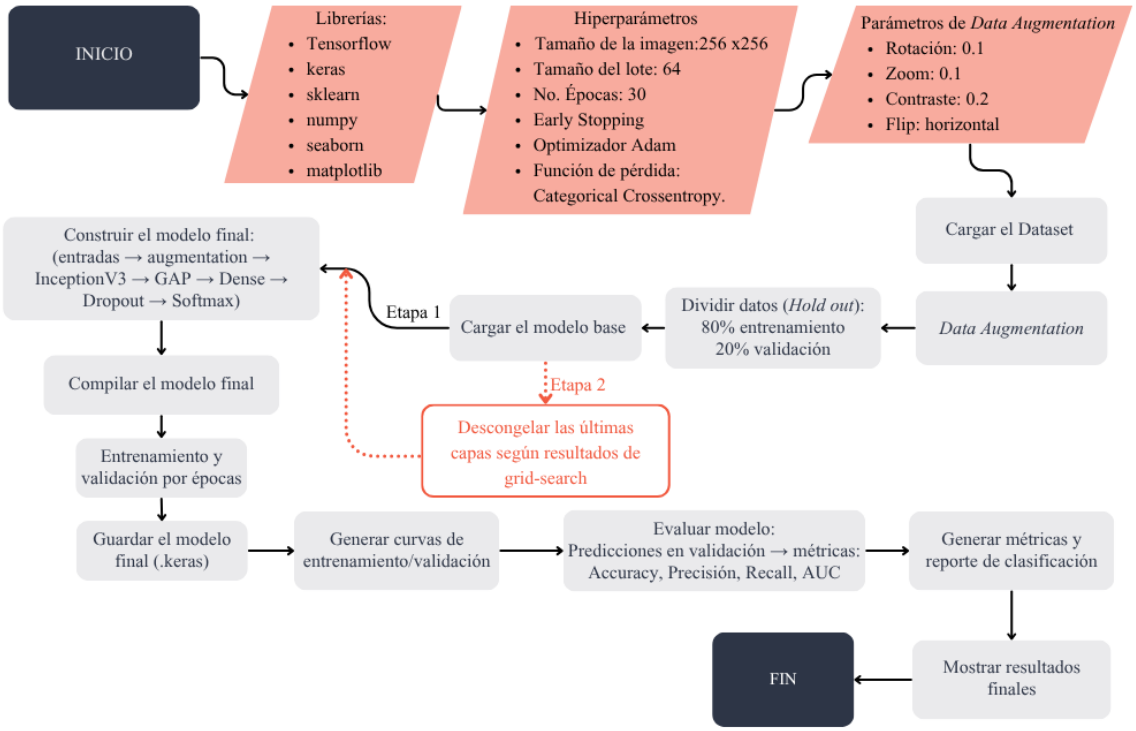
Integrar métodos de interpretabilidad (Grad-CAM o LIME) para comprender las regiones de atención del modelo y respaldar su aplicación segura en contextos médicos. De manera complementaria, el modelo podría beneficiarse de estrategias de mitigación del desequilibrio, tales como la asignación de pesos por clase (*class weighting*), el uso de funciones de pérdida adaptativas como *focal loss* [Lin et al., 2017] o la aplicación de técnicas de sobre-muestreo controlado.

Finalmente, implementar una prueba piloto en entornos hospitalarios reales que permitiría evaluar la usabilidad, seguridad y aceptación del sistema por parte del personal médico.

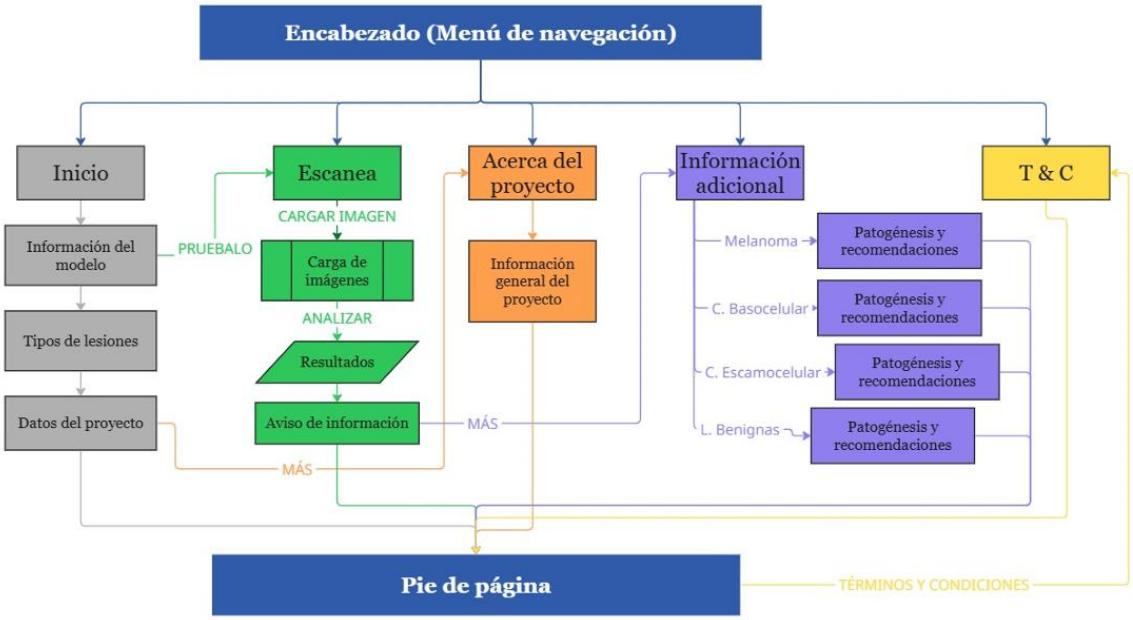
8. ANEXOS

A. ANEXO: DIAGRAMA DE FLUJO DEL MODELO

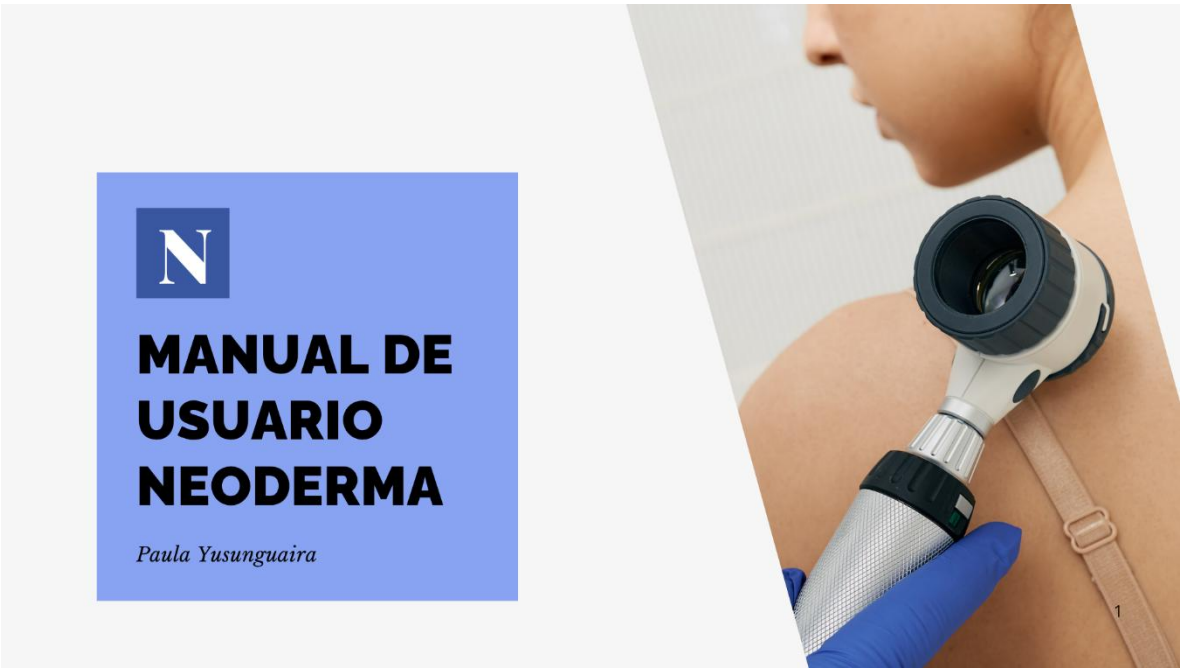
Diagrama de flujo de código del modelo

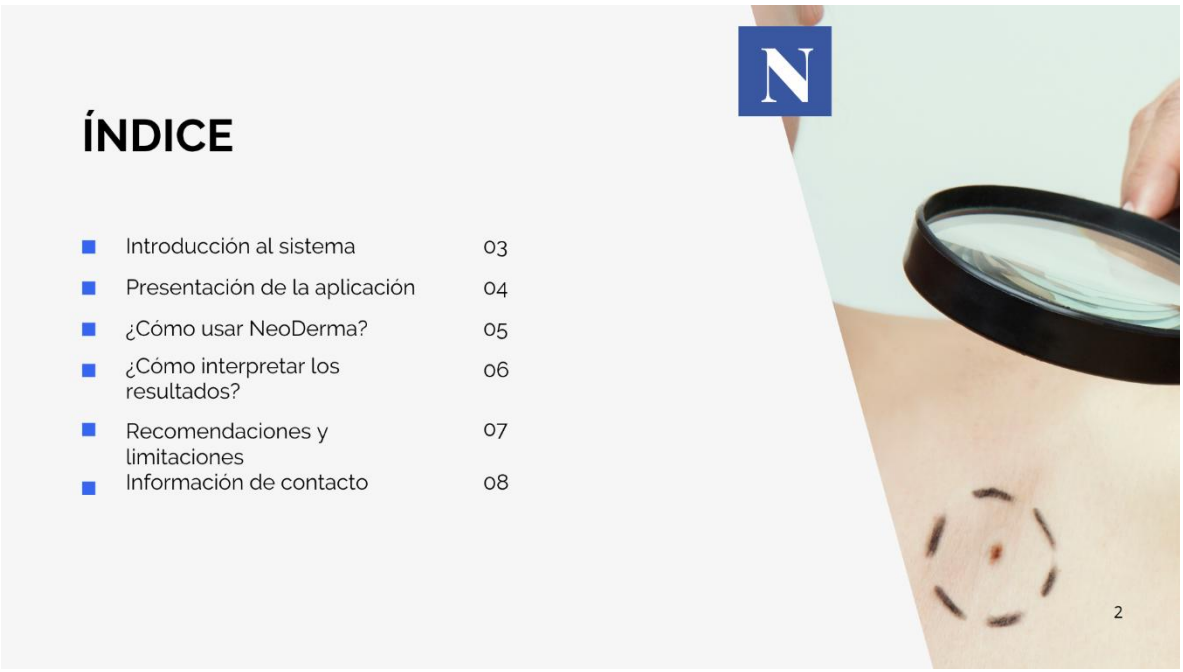


B. ANEXO: DIAGRAMA DE FLUJO DE NAVEGACIÓN



C. ANEXO: MANUAL DE USUARIO



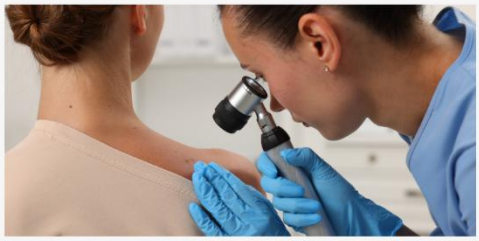


INTRODUCCIÓN

NeoDerma es una herramienta digital diseñada para apoyar el tamizaje temprano del cáncer de piel mediante la clasificación automática de imágenes dermatoscópicas.

La interfaz permite cargar una imagen, procesarla con el modelo de inteligencia artificial y visualizar la probabilidad asociada a cada categoría diagnóstica.

Diseñado para evaluar la viabilidad de implementación de estos modelos entrenados con bases de datos internacionales en nuestra población colombiana con sus variados fototipos de piel.



3

PRESENTACIÓN DE LA APLICACIÓN (PESTAÑAS DE LA INTERFAZ)

INICIO

Presentación del sistema NeoDerma y acceso directo a la sección de análisis

ESCANEA

Permite el análisis de la imagen (Véase en la página 5)

ACERCA DEL PROYECTO

Información sobre objetivos, alcance y fundamentos del sistema.

INFORMACIÓN ADICIONAL

Incluye recomendaciones, advertencias y consideraciones de uso clínico.

T & C

Explica las limitaciones de uso y políticas de manejo de datos.

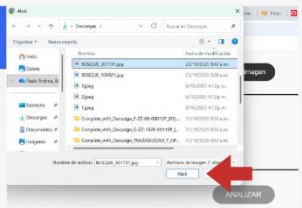
4

¿CÓMO USAR NEODERMA?

1

En la pestaña '**Escanea**' se encontrará esta presentación. Presione '**Cargar imagen**' para comenzar.

2



Seleccione la imagen que desee cargar y presione **'Abrir'**

3

Presione **'Analizar'** y obtendrá los resultados de predicción en PDF

4

Lea su informe (para interpretar resultados vea la página 6), si desea más información sobre el tipo de lesión arrojada dirijase a la pestaña **'Información adicional'**

5

¿CÓMO INTERPRETAR LOS RESULTADOS?

PASO NÚMERO 1
Revise la clasificación/predicción final del modelo (Melanoma, Carcinoma basocelular, Carcinoma escamocelular, lesión benigna)

PASO NÚMERO 2
Identifique el porcentaje de probabilidad arrojado por la interfaz, agende cita con su médico lo más pronto posible en caso de lesiones malignas.

PASO NÚMERO 3
Mantenga la calma, lea las recomendaciones dadas por el sistema. Recuerde que no es una herramienta diagnóstica

PASO NÚMERO 4
Si la probabilidad no supera el 50% los resultados no son confiables, el modelo no está seguro de su predicción por ende no se confie y acuda un profesional.

NOTA
Este resultado no es un diagnóstico definitivo ante cualquier cambio conducirse a una evaluación médica.

6

RECOMENDACIONES DE USO

- **Utilice imágenes tomadas con dermatoscopio.**
- **Evite imágenes borrosas, con sombras o cuerpos extraños.**
- **No manipule la imagen antes de cargarla (no recortar ni alterar).**
- **Use el sistema como apoyo, nunca como herramienta diagnóstica.**

LIMITACIONES

- No reemplaza la valoración profesional de un dermatólogo.
- El rendimiento puede verse afectado por imágenes de muy baja calidad.
- El modelo fue entrenado principalmente con ISIC y validado con muestras colombianas; resultados pueden variar con otros tipos de dispositivos o condiciones de iluminación.

7

NeoDerma

INFORMACIÓN DE CONTACTO

Para reportar problemas

CORREO

pyusunguaira@unbosque.edu.co

8

BIBLIOGRAFÍA

ABINI, M. A.; SATHYA PRIYA, S. S. "Detection and Classification of Diabetic Retinopathy Using Modified Inception V3". *International Journal Bioautomation*, vol. 29, n.º 1, 2025.

ALARCÓN, J. A. L.; TERÁN, C. E. E. *Desarrollo de un algoritmo basado en redes neuronales para la identificación automática del melanoma mediante el análisis de imágenes*. Quito: Pontificia Universidad Católica del Ecuador, Facultad de Ingeniería, 2025.

AMERICAN CANCER SOCIETY. "Pruebas para el cáncer de piel tipo melanoma". *Cancer.org*, 14 de agosto de 2019. Disponible en: https://www.cancer.org/es/cancer/tipos/cancer-de-piel-tipo-melanoma/deteccion-diagnostico-clasificacion-por-etapas/como-se-diagnostica.html [Consulta: 18 de septiembre de 2024].

BARRERA, S. M. R. "Melanoma en Colombia: prevalencia y diagnóstico en 2021". *ConsultorSalud*, 27 de mayo de 2022. Disponible en: https://consultorsalud.com/melanoma-en-colombia-prevalencia-diagnos [Consulta: 18 de octubre de 2025].

BELETE, D. M.; HUCHAIAH, M. D. "Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results". *International Journal of Computers and Applications*, vol. 44, n.º 9, p. 875–886, 2022.

BRINKER, T. J.; HEKLER, A.; ENK, A. H.; BERKING, C.; et al. "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task". *European Journal of Cancer*, vol. 113, p. 47–54, 2019. DOI: https://doi.org/10.1016/j.ejca.2019.04.001

BUHRMESTER, V.; MUENCH, D.; ARNS, R.; et al. "Analysis of Bias in Deep Neural Networks for Medical Image Classification". *Artificial Intelligence in Medicine*, vol. 121, 2021.

CANCER CENTER MOFFITT. "¿Cuáles son los síntomas del melanoma que se ha extendido?". *Moffitt* (sitio web), 2023. Disponible en: https://www.moffitt.org/es/cancers/melanoma/faqs/what-are-the-symptoms-of-melanoma-that-has-spread/ [Consulta: 20 de octubre de 2025].

CASARIEGO, Z. J.; BAUDO, J. E. "Trabajo de revisión: melanoma". *Avances en Periodoncia e Implantología Oral*, vol. 16, n.º 3, p. 157–177, 2004.

CLÍNICA UNIVERSIDAD DE NAVARRA. "Relación señal a ruido (SNR)". *Diccionario Médico CUN*, 2023. Disponible en: https://www.cun.es/diccionario-medico/terminos/senal-ruido [Consulta: 18 de octubre de 2025]. [Collazos, 2021] Collazos, C. (2021, August 5). Datos | Cáncer de Piel - Liga Colombiana contra el Cáncer. Liga

Colombiana contra el Cáncer. https://www.ligacancercolombia.org/educacion/datos-cancer-de-piel/

CHOLLET, F. Xception: deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, p. 1251–1258. IEEE. Disponible en: https://doi.org/10.1109/CVPR.2017.195

CRUZ GUEVARA, R. S.; DELRIEUX, C. A. Aplicación de redes neuronales densas y convolucionales para detección de COVID_19 en imágenes de rayos X. Conectividad, 2023, vol. 4, no 2, p. 19-32.

DINNES, J.; DEEKS, J. J.; CHALMERS, J. R.; et al. "Visual inspection and dermoscopy for diagnosing cutaneous melanoma in adults". *Cochrane Database of Systematic Reviews*, n.º 12, 2018. [Espasa, 2022] Espasa Rosell, J. (2022). Deep learning en la detección de lesiones cutáneas malignas (Bachelor's thesis, Universitat Politècnica de Catalunya).

FLÓREZ-MORALES, I.; et al. "Reporte de la experiencia de la Sección de Dermatología en la 5ª campaña de tamizaje de cáncer de piel". *Iatreia*, vol. 31, n.º 4, p. 362–370, 2018.

GARCÍA ABAD, J. "Aplicación de técnicas híbridas de resampling para el balanceo de datos en clasificación supervisada". *Revista de Ingeniería y Tecnología*, vol. 15, n.º 2, 2021.

GEEKSFORGEEKS. "CLAHE Histogram Equalization OPENCV". *GeeksforGeeks* (sitio web), 19 de mayo de 2023. Disponible en: https://www.geeksforgeeks.org/clahe-histogram-eqalization-opencv/ [Consulta: 20 de mayo de 2025].

HUANG, G., LIU, Z., VAN DER MAATEN, L. Y WEINBERGER, K. Q. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, p. 4700–4708. IEEE. Disponible en: https://doi.org/10.1109/CVPR.2017.243

HUIX PÉREZ, A. *Clasificación de lesiones cutáneas mediante inteligencia artificial*. Barcelona: Universitat Politècnica de Catalunya, 2024.

INSTITUTO NACIONAL DE CANCEROLOGÍA COLOMBIA. *Dermatología Oncológica: Generalidades*. Bogotá: Instituto Nacional de Cancerología, 2021. Disponible en: https://www.cancer.gov.co/portafolio-1/unidades-funcionales-1/dermatologia-oncologica/generalidades [Consulta: 11 de febrero de 2025].

INSTRUMENTS & DATA TOOLS. "Unsharp masking with Python and OpenCV". *idtools.com.au*, 5 de mayo de 2021. Disponible en: https://www.idtools.com.au/unsharp-masking-with-python-and-opencv/ [Consulta: 22 de junio de 2025].

ISIC – INTERNATIONAL SKIN IMAGING COLLABORATION. *ISIC Archive*. 2024. Disponible en: https://www.isic-archive.com [Consulta: 18 de octubre de 2025].

JARAMILLO ARBOLEDA, A.; SÁNCHEZ ZAPATA, M. J.; RUEDA JAIME, L. J.; MORALES-FORERO, A.; BASSETTO, S. “Conjunto de pruebas BOSQUE”. *Harvard Dataverse*, V1, 2025. DOI: https://doi.org/10.7910/DVN/AQEPIN

JUAN, R. Q.; MARIO, C. M. “Redes neuronales artificiales para el procesamiento de imágenes: una revisión de la última década”. *Revista de Ingeniería Eléctrica, Electrónica y Computación (RIEE&C)*, vol. 9, n.º 1, 7–16, 2011.

LEE, T.; NG, V.; GALLAGHER, R.; COLDMAN, A.; McLEAN, D. “Dullrazor: A software approach to hair removal from images”. *Computers in Biology and Medicine*, vol. 27, n.º 6, 533–543, 1997.

MAHARANA, K.; MONDAL, S.; NEMADE, B. “A review: Data preprocessing methods for medical image classification”. *Biomedical Signal Processing and Control*, vol. 71, 2022.

MASCARENHAS, S.; AGARWAL, M. “A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification”. En: *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, vol. 1, p. 96–99, IEEE, noviembre 2021.

MATHWORKS. “¿Qué son las redes neuronales convolucionales?” *MathWorks* (sitio web), 2022. Disponible en: https://es.mathworks.com/discovery/convolutional-neural-network.html [Consulta: 20 de octubre de 2025].

MÉNDEZ MATTHEY, V. E. “Queratosis actínica”. *Revista Médica Herediana*, vol. 27, n.º 1, p. 67, 2016.

MÉNDEZ MATTHEY, V. E. “Queratosis seborreica”. *Revista Médica Herediana*, vol. 28, n.º 4, p. 266, 2017.

MORALES-FORERO, A.; RUEDA, L. J.; HERRERA, R.; BASSETTO, S.; COATANEA, E. “Predictive Representativity: Uncovering Racial Bias in AI-based Skin Cancer Detection”. *arXiv* preprint arXiv:2507.14176, 2025.

MORALES KAMILAH, A.; LEPE, D. R.; DÍAZ, J. F.. Redes Neuronales Convolucionales para Detección de Carcinoma Basocelular (Cáncer de Piel) por Medio de Imágenes para Pacientes Adultos en la Región Facial. *Ciencia Latina Revista Científica Multidisciplinar*, 2024, vol. 8, no 4, p. 582-599.

MUMUNI, A.; MUMUNI, F. “Data augmentation: A comprehensive survey of modern approaches”. *Array*, vol. 16, art. 100258, 2022.

MYERS, D. J.; FILLMAN, E. P. “Dermatofibroma”. En: *StatPearls* [Internet]. StatPearls Publishing; 2024. Disponible en: https://www.ncbi.nlm.nih.gov/books/NBK538317/ [Consulta: 20 de octubre de 2025].

NGUYEN, V. D.; BUI, N. D.; DO, H. K. “Skin Lesion Classification on Imbalanced Data Using Deep Learning with Soft Attention”. *Sensors*, vol. 22, n.º 19, art. 7530, 2022. DOI: https://doi.org/10.3390/s22197530.

NOGUEIRA, V.; LORENZETTI, M.; PINHO, M.; RODRIGUES, J. “Skin cancer diagnosis: The importance of screening and early detection”. *International Journal of Dermatology*, vol. 60, n.º 12, p. 1445–1453, 2021. DOI: https://doi.org/10.1111/ijd.15501

OSPIÑA, J. P.; MONTES, C.; LÓPEZ, N. “Epidemiología del melanoma en Colombia: tendencias recientes”. *Revista Colombiana de Oncología*, vol. 25, n.º 1, 2021.

PATINO-PEREZ, Darwin; ARMIJOS-VALAREZO, Luis; CHOEZ-ACOSTA, Luis y BURGOS-ROBALINO, Freddy. REDES NEURONALES CONVOLUCIONALES PARA DETECCIÓN DE RETINOPATÍA DIABETICA. Ingenius [online]. 2025, n.33 [citado 2025-11-10], pp.91-101. Disponible en: <http://scielo.senescyt.gob.ec/scielo.php?script=sci_arttext&pid=S1390-860X2025000100091&lng=es&nrm=iso>. ISSN 1390-860X. https://doi.org/10.17163/ings.n33.2025.08.

PARDO-ZAMUDIO, A. C.; et al. “Prúrigo actínico en un centro dermatológico de referencia en Colombia: 108 casos”. *Biomédica*, 2020.

PMU ONLINE TRAINING. “Escala de Fitzpatrick y elección correcta de pigmento”. *PMU Online Training* (sitio web), 2025. Disponible en: https://pmuonlinetraining.com/topic/escala-de-fitzpatrick-y-eleccion-correcta-de-pigmento-2-2/ [Consulta: 20 de octubre de 2025].

PUTRI, A. N. P.; NOOR, M. N. M. M.; DARWIS, H.; HAYATI, L. N.; AS’AD, I. “A Comparative Study on Efficacy of CNN VGG-16, DenseNet121, ResNet50V2, And EfficientNetB0 in Toraja Carving Classification”. *Indonesian Journal of Data and Science*, vol. 6, n.º 1, p. 122–131, 2025.

QIANG, Z.; HE, L.; DAI, F. “Identification of plant leaf diseases based on Inception V3 transfer learning and fine-tuning”. En: *International Conference on Smart City and Informatization*. Singapore: Springer, 2019. p. 118–127.

RESTREPO VÉLEZ, J. C.; et al. “Conocimientos y prácticas frente a la exposición solar y tamizaje de cáncer de piel”. *Revista de la Asociación Colombiana de Dermatología*, 2001.

SAM, S. M.; KAMARDIN, K.; SJARIF, N. N. A.; MOHAMED, N. “Offline signature verification using deep learning convolutional neural network architectures GoogLeNet Inception-V1 and Inception-V3”. *Procedia Computer Science*, vol. 161, 475–483, 2019.

SERRANO, A. V. O.; BERNAL, L. “Tratamiento del melanoma maligno en Colombia: resultados encuesta ACHO y proyecto grupo colombo-español del tratamiento

multidisciplinario del melanoma". *Revista Colombiana de Hematología y Oncología*, vol. 6, n.º 1, 33–37, 2019.

SECRETARÍA DE SALUD. "Falta de prevención y detección tardía ocasiona 30% de fallecimientos en pacientes con melanoma". *gob.mx*, 24 de mayo de 2024. Disponible en: https://www.gob.mx/salud/prensa/196-falta-de-prevencion-y-deteccion-tardia-ocasiona-30-de-fallecimientos-en-pacientes-con-melanoma [Consulta: 18 de octubre de 2025].

SHANTHI, T.; SABEENIAN, R. S. "Modified AlexNet architecture for classification of diabetic retinopathy images". *Computers & Electrical Engineering*, vol. 76, 56–64, 2019.

SKIN CANCER FINANCIAL ASSISTANCE. "Skin cancer financial assistance". *Help Hope Live*, 26 de octubre de 2023. Disponible en: https://helphopelive.org/skin-cancer-financial-assistance/ [Consulta: 18 de octubre de 2025].

THE SKIN CANCER FOUNDATION. "Datos y estadísticas sobre el cáncer de piel". *The Skin Cancer Foundation*, 6 de febrero de 2024. Disponible en: https://www.skincancer.org/es/skin-cancer-information/skincancerfacts [Consulta: 18 de octubre de 2025].

URIBE, P.; NOVA, J. A.; COLMENARES, C.; PALMA, L. F.; GU-QUIÑONES, S. R. "Características del melanoma cutáneo en dos instituciones de Bogotá, Colombia: Análisis 2012–2016". *Revista Colombiana de Cancerología*, vol. 25, n.º 4, p. 188–195, 2021.

VIKNESH, C. K.; KUMAR, P. N.; SEETHARAMAN, R.; ANITHA, D. "Detection and classification of melanoma skin cancer using image processing technique". *Diagnostics*, vol. 13, n.º 21, art. 3313, 2023.

VINCENT, L. "Morphological area openings and closings for grey-scale images". En: *Shape in Picture: Mathematical Description of Shape in Grey-level Images*, Springer Berlin Heidelberg, 1994. p. 197–208.

VRBANČIČ, G.; PODGORELEC, V. "Transfer learning with adaptive fine-tuning". *IEEE Access*, vol. 8, p. 196197–196211, 2020.

VELÁSQUEZ, P. "Cáncer de piel: ¿cómo está Colombia en esta materia?". *ConsultorSalud*, 18 de junio de 2020. Disponible en: https://consultorsalud.com/cancer-de-piel-como-esta-colombia-en-esta-materia/ [Consulta: 18 de octubre de 2025].

VELÁSQUEZ, J. *GitHub - BlueDokk/Dullrazor-algorithm: Pre-processing technique called DullRazor for the detection and removal of hairs on dermoscopic images*. GitHub, 2020. Disponible en: https://github.com/BlueDokk/Dullrazor-algorithm [Consulta: 18 de octubre de 2025].

WORLD HEALTH ORGANIZATION (WHO). “Cáncer”. *WHO Fact Sheets*, 3 de febrero de 2022. Disponible en: https://www.who.int/es/news-room/fact-sheets/detail/cancer [Consulta: 18 de octubre de 2025].

ZHANG, Z. “Improved Adam optimizer for deep neural networks”. En: *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, 2018. p. 1–2. IEEE.