

Desarrollo de un sistema de apoyo al diagnóstico de cáncer de piel basado en modelos híbridos  
de aprendizaje profundo integrados en una aplicación web

Autores

*Diego Fernando Fonseca Miranda*

*Donovan David Torres Vahos*

*Santiago Andrés Chain Santos*

Trabajo de Grado para Optar el Título de Ingeniero de Sistemas e Informática

Director

Dra. Lola Xiomara Bautista Rozo

*Ph.D. Automática, Procesamiento de Señales e Imágenes Digitales*

Codirector

Zully Johanna Ballesteros Zárate

*Médica especialista en dermatología clínica*

Universidad Industrial de Santander  
Facultad de Ingenierías Fisicomecánicas  
Escuela de Ingeniería de Sistemas e Informática  
Bucaramanga

2026

### **Dedicatoria**

A mi familia, por su amor incondicional, sus sacrificios silenciosos y su apoyo constante a lo largo de este camino. Gracias por inculcarme los valores y la fortaleza necesarios para alcanzar este logro.

A mi novia, por su amor, comprensión y paciencia durante cada etapa de este proceso, por acompañarme en los momentos de dificultad y por ser un apoyo inquebrantable que me impulsó a seguir adelante.

A su familia, especialmente a sus padres, por su generosidad, su respaldo sincero y por acogerme con aprecio, brindándome apoyo y motivación a lo largo de este proceso. Su confianza y acompañamiento han sido también parte fundamental de este camino.

Este logro lleva consigo una parte de ustedes.

***Diego Fernando Fonseca Miranda***

Dedico este logro, antes que nada, a mi mamá, por ser el pilar más firme en mi vida, por su amor incondicional, por cada esfuerzo silencioso y por enseñarme a seguir adelante incluso en los momentos más difíciles. Todo lo que soy y lo que estoy construyendo tiene una raíz profunda en su ejemplo.

A mi hermana, por ser parte de mi historia de una forma tan especial, por los momentos compartidos, por enseñarme sobre el cuidado, la paciencia y el amor en medio de las dificultades. Este camino también es por y para nosotros.

A mis gatos, que con su presencia tranquila y su cariño silencioso han sido compañía en largas noches, en momentos de estrés y en instantes donde solo necesitaba sentir calma. Han sido más importantes de lo que muchos imaginaron.

A mis amigos, quienes han estado en las buenas y en las no tan buenas, por las risas, el apoyo, las conversaciones y por recordarme que no estoy solo en este camino.

A mis profesores, por compartir su conocimiento, por su paciencia y por dejar huellas que van más allá de lo académico.

Y a mi psicóloga, por ayudarme a entenderme, a reconstruirme y a avanzar. Por acompañarme en procesos internos que han sido clave para llegar hasta aquí.

Este logro no es solo mío, es de todos ustedes.

*Donovan David Torres Vahos*

A mi madre Janeth, por ser el pilar de mi vida, mi apoyo incondicional en los momentos más difíciles y darme la fuerza que me impulsa a seguir adelante y ser cada día una mejor persona. Gracias por enseñarme con tu ejemplo el valor de la dedicación, constancia y amor por lo que se hace.

A mi padre Humberto, por ser mi guía en cada proceso y por orientarme con sabiduría durante todo el camino de la vida. Gracias por siempre estar ahí cuando más te necesitaba y por enseñarme a mejorar, aprender de mis errores y a nunca dejar de crecer.

A mi familia, por arroparme con cariño y confianza en todo lo que me he propuesto. Gracias por estar presentes y creer en mí desde mi primer día.

*Santiago Andrés Chaín Santos*

## Tabla de Contenido

	<b>Pág.</b>
Introducción .....	13
1. Objetivos .....	16
1.1 Objetivo General .....	16
1.2 Objetivos Específicos.....	16
2. Marco Referencial.....	17
2.1. Cáncer de piel .....	17
2.2. Aprendizaje profundo aplicado al análisis de imágenes médicas .....	19
2.3. Redes Neuronales Convolucionales ( <i>CNNs</i> ) .....	20
2.4. Vision Transformers ( <i>ViTs</i> ) .....	22
2.5. Modelos híbridos <i>CNN + ViT</i> .....	23
2.6. Aprendizaje profundo aplicado a la clasificación de lesiones cutáneas .....	25
2.7. Diagnóstico asistido por IA en Colombia.....	26
3. Método propuesto .....	27
3.1. Preparación de datos .....	27
3.1.1. Dataset ISIC 2019.....	27
3.1.2. Construcción, integración y alteración de dataset para entrenamiento.....	30
3.2. Métricas de evaluación .....	32
3.2.1. Selección del umbral óptimo .....	34
3.2.2. Intervalos de confianza bootstrap .....	35
3.2.3. Calibración probabilística.....	35
3.3. Modelo <i>CNN</i> .....	36
3.3.1. Selección de arquitectura e hiperparámetros .....	36
3.3.2. Entrenamiento .....	38
3.3.3. Resultados.....	40
3.4. Modelo <i>VIT</i> .....	43
3.4.1. Selección de arquitectura e hiperparámetros .....	43
3.4.2. Entrenamiento .....	44
3.4.3. Resultados.....	46
3.5. Modelo Híbrido <i>CNN + VIT</i> .....	49
3.5.1. Motivación para la fusión de modelos.....	49
3.5.2. Arquitectura del módulo de fusión .....	50
3.5.3. Entrenamiento .....	51
3.5.4. Resultados.....	53
3.5.4.1 Resultados del entrenamiento. ....	53
3.5.4.2 Resultados de métricas e intervalo de confianza. ....	54

3.5.4.3. Resultados absolutos.....	55
3.5.4.4. Attention gate.....	56
3.5.4.5. Calibración Probabilística.....	59
3.5.4.6. Estudio de ablación.....	60
3.6. Desarrollo del prototipo web .....	61
3.6.1 Requerimientos funcionales.....	62
3.6.2 Casos de uso.....	63
3.6.3 Diseño del sistema .....	63
3.6.4 Diseño de la base de datos .....	64
3.6.5 Diseño del backend.....	66
3.6.6 Diseño del frontend.....	70
3.6.7 Tecnologías y arquitectura de la interfaz.....	73
3.6.8 Planificación de módulos y estructura de la interfaz .....	75
3.6.9 Flujo de interacción del especialista .....	76
3.6.10 Responsividad y diseño para dispositivos móviles.....	78
3.6.11 Metodología de desarrollo .....	78
3.6.12 Pruebas técnicas y de usabilidad.....	80
3.6.12.1 Pruebas técnicas.....	80
3.6.12.2 Pruebas de usabilidad.....	82
3.6.13 Evaluación y retroalimentación .....	83
3.7. Discusión de resultados.....	84
3.7.1 Resultados de Funcionalidad de la Aplicación .....	85
4. Conclusiones .....	87
5. Trabajo futuro .....	88
Referencias Bibliográficas .....	90
Apéndices.....	96

**Lista de Tablas**

	<b>Pág.</b>
Tabla 1. Distribución de imágenes por categoría binaria en el corpus ISIC 2019.....	30
Tabla 2. Particiones del corpus ISIC.....	31
Tabla 3. Métricas de evaluación clínica empleadas.....	34
Tabla 4. Experimentación Modelo CNN .....	37
Tabla 5. Configuración final Modelo CNN .....	39
Tabla 6. Reporte de clasificación modelo CNN .....	41
Tabla 7. Experimentación Modelo ViT .....	44
Tabla 8. Configuración final Modelo ViT .....	45
Tabla 9. Reporte de clasificación modelo ViT .....	47
Tabla 10. Componentes y parámetros del modelo híbrido .....	51
Tabla 11. Experimentación Modelo Híbrido .....	52
Tabla 12. Resultados de valores obtenidos vs IC .....	54
Tabla 13. Pacientes .....	64
Tabla 14. Diagnóstico .....	65
Tabla 15. Módulos del frontend y requerimientos funcionales asociados.....	75
Tabla 16. Pruebas técnicas del prototipo web.....	80
Tabla 17. Resultados pruebas técnicas del prototipo web .....	86
Tabla 18. Requisitos de acceso a la plataforma .....	107
Tabla 19. Estructura general de la plataforma .....	108

## Lista de Figuras

	<b>Pág.</b>
Figura 1: .....	18
<i>Esquema de los principales tipos de cáncer de piel.</i> .....	18
Figura 2. ....	21
<i>Ejemplo de estructura básica de una red CNN</i> .....	21
Figura 3. ....	23
<i>Arquitectura del Visión Transformer y estructura detallada del bloque codificador</i> .....	23
Figura 4. ....	24
<i>Esquema del modelo híbrido Transformer-CNN para predicción.</i> .....	24
Figura 5. ....	29
<i>Ejemplos de las diversas categorías en el dataset ISIC 2019.</i> .....	29
Figura 6. ....	40
<i>Curvas de entrenamiento — CNN EfficientNet-B0</i> .....	40
Figura 7. ....	42
<i>Matriz de confusión - Modelo CNN</i> .....	42
Figura 8. ....	46
<i>Curvas de entrenamiento — ViT DeiT-B/16</i> .....	46
Figura 9. ....	48
<i>Matriz de confusión - Modelo ViT</i> .....	48
Figura 10. ....	53
<i>Curvas de entrenamiento — Modelo Híbrido CNN + ViT (Exp. 2, dropout = 0,30)</i> .....	53
Figura 11. ....	55
<i>Evaluación clínica — Modelo Híbrido CNN + ViT</i> .....	55
Figura 12. ....	57
<i>Peso medio de atención por rama según clase diagnóstica</i> .....	57
Figura 13. ....	58
<i>Peso individual de atención por rama</i> .....	58
Figura 14. ....	59
<i>Calibración probabilística — Modelo Híbrido CNN + ViT</i> .....	59
Figura 15. ....	60
<i>Estudio de ablación – contribución de cada rama al AUC-ROC del modelo híbrido.</i> .....	60
Figura 16. ....	63
<i>Diagrama de Casos de Uso</i> .....	63
Figura 17. ....	66
<i>Esquema visual de la base de datos</i> .....	66
Figura 18. ....	69

<i>Arquitectura del backend</i> .....	69
Figura 19. ....	70
<i>Diagrama de secuencia</i> .....	70
Figura 20. ....	71
<i>Mockups del sistema</i> .....	71
Figura 21. ....	74
<i>Arquitectura y Tecnologías del Frontend</i> .....	74
Figura 22. ....	77
<i>Diagrama de Actividades del Flujo de la Aplicación</i> .....	77
Figura 23. ....	115
<i>Respuestas de la Sección 1 del Formulario de Evaluación de Usabilidad de DermaScan</i> .....	115
Figura 24. ....	117
<i>Respuestas de la Sección 2 del Formulario de Evaluación de Usabilidad de DermaScan (Escala SUS)</i> . ....	117
Figura 25. ....	124
<i>Respuestas de la Sección 4 del Formulario de Evaluación de Usabilidad de DermaScan</i> .....	124

**Lista de Apéndices**

APÉNDICE A. TABLAS DE REQUERIMIENTOS FUNCIONALES Y NO FUNCIONALES Y CASOS DE USO .....	96
APÉNDICE B . MANUAL DE USUARIO – DERMASCAN .....	107
APÉNDICE C. RESULTADOS DE LAS PRUEBAS DE USABILIDAD .....	114

## Resumen

**Título:** Desarrollo de un sistema de apoyo al diagnóstico de cáncer de piel basado en modelos híbridos de aprendizaje profundo integrados en una aplicación web.<sup>1\*</sup>

**Autores:** Diego Fernando Fonseca Miranda, Donovan David Torres Vahos, Santiago Andrés Chain Santos <sup>2\*\*</sup>

**Palabras Clave:** Cáncer de piel, Aprendizaje profundo, Modelos híbridos, Aplicación web, Diagnóstico asistido, Redes neuronales convolucionales.

### Descripción:

El cáncer de piel es una de las neoplasias más frecuentes a nivel mundial, cuya incidencia continúa en aumento, haciendo de la detección temprana un factor crucial para mejorar el pronóstico y reducir la mortalidad. Históricamente, el diagnóstico clínico depende de la inspección visual y la dermatoscopia, métodos que, aunque efectivos, están sujetos a la variabilidad inter-observador y a la disponibilidad de especialistas calificados. Esta situación resalta la necesidad de integrar tecnologías automatizadas que asistan de manera objetiva la toma de decisiones clínicas.

Para abordar esta problemática, este trabajo de grado presenta el desarrollo de un sistema de apoyo al diagnóstico capaz de clasificar lesiones cutáneas sospechosas de malignidad. La solución se fundamenta en la implementación de un modelo híbrido de aprendizaje profundo que combina el poder de extracción de características locales de las Redes Neuronales Convolucionales (CNN) con la capacidad de modelar dependencias globales de largo alcance proporcionada por los Vision Transformers (ViT). Este enfoque sinérgico permite analizar tanto patrones morfológicos sutiles como la estructura general de la lesión, mejorando la precisión diagnóstica frente a arquitecturas individuales. El sistema se entrena y evalúa utilizando imágenes dermatoscópicas provenientes de repositorios de acceso abierto, aplicando rigurosas métricas de clasificación médica.

Con el fin de garantizar la accesibilidad y aplicabilidad clínica, el modelo de inteligencia artificial se integra en un prototipo funcional de aplicación web bajo una arquitectura cliente-servidor. Mediante esta plataforma web, el personal médico puede interactuar de forma intuitiva con el sistema desde cualquier navegador, facilitando el análisis automatizado sin requerir instalaciones locales. Esta herramienta tecnológica busca optimizar el diagnóstico dermatológico, facilitar el acceso a servicios especializados y contribuir activamente a las estrategias de salud digital vigentes.

---

<sup>1\*</sup> Trabajo de Grado

<sup>2\*\*</sup> Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Dra. Lola Xiomara Bautista Rozo. Ph.D. Automática, Procesamiento de Señales e Imágenes Digitales Codirector: Zully Johanna Ballesteros Zárate. Médica especialista en dermatología clínica

### Abstract

**Title:** Development of a support system for the diagnosis of skin cancer based on hybrid deep learning models integrated into a web application.

**Authors:** Diego Fernando Fonseca Miranda, Donovan David Torres Vahos, Santiago Andrés<sup>3\*</sup>

**Keywords:** Skin cancer, Deep learning, Hybrid models, Web application, Computer-aided diagnosis, Convolutional neural networks.<sup>4\*\*</sup>

### Description:

Skin cancer is one of the most frequent neoplasms worldwide, with an incidence that continues to rise, making early detection a crucial factor in improving prognosis and reducing mortality. Historically, clinical diagnosis has relied on visual inspection and dermoscopy, methods that, although effective, are subject to inter-observer variability and the availability of qualified specialists. This situation highlights the need to integrate automated technologies that objectively assist in clinical decision-making.

To address this issue, this degree project presents the development of a diagnostic support system capable of classifying skin lesions suspected of malignancy. The solution is based on the implementation of a hybrid deep learning model that combines the local feature extraction power of Convolutional Neural Networks (CNN) with the ability to model long-range global dependencies provided by Vision Transformers (ViT). This synergistic approach allows for the analysis of both subtle morphological patterns and the overall structure of the lesion, improving diagnostic accuracy compared to individual architectures. The system is trained and evaluated using dermoscopic images from open-access repositories, applying rigorous medical classification metrics.

To ensure accessibility and clinical applicability, the artificial intelligence model is integrated into a functional web application prototype under a client-server architecture. Through this web platform, medical personnel can interact intuitively with the system from any browser, facilitating automated analysis without requiring local installations. This technological tool seeks to optimize dermatological diagnosis, facilitate access to specialized services, and actively contribute to current digital health strategies.

---

<sup>3</sup> \*Degree Work

<sup>4\*\*</sup> Faculty of Physical-Mechanical Engineering. School of Systems and Computer Engineering. Director: Dr. Lola Xiomara Bautista Roza, Ph.D. in Automation, Signal Processing, and Digital Images. Co-director: Zully Johanna Ballesteros Zárate, MD, Specialist in Clinical Dermatology

## Introducción

El cáncer de piel es una de las neoplasias con mayor incidencia a nivel mundial y su tendencia sigue en aumento. Para el año 2022, la Organización Mundial de la Salud estimó aproximadamente 331 722 nuevos casos de melanoma y 58 667 muertes asociadas en todo el mundo [1]. Si bien las cifras son llamativas por sí solas, lo que convierte a esta enfermedad en una prioridad de salud pública es la enorme brecha que existe en el pronóstico según el momento del diagnóstico: cuando la lesión se detecta en estudios tempranos, la supervivencia a cinco años supera el 99 %; en estudios avanzados, ese porcentaje cae a aproximadamente el 35 % [2]. Esto hace de la detección temprana no solo un objetivo clínico deseable, sino determinante para la vida del paciente.

Históricamente, el diagnóstico de las lesiones cutáneas sospechosas ha recaído sobre la inspección visual del especialista, apoyada progresivamente por la dermatoscopia. Esta última técnica, también denominada microscopía de epiluminiscencia, incrementó la sensibilidad diagnóstica del melanoma de aproximadamente 76,9 % a 85,7 % al permitir la visualización de estructuras subepidérmicas invisibles al ojo desnudo [3]. Con todo, la interpretación de las imágenes dermatoscópicas sigue siendo un proceso subjetivo y altamente dependiente de la experiencia del observador, con variabilidad documentada incluso entre especialistas expertos [1]. En contextos con alta demanda asistencial o con escasez de dermatólogos —como ocurre en gran parte del territorio colombiano y latinoamericano— esta limitación se amplifica y genera desigualdades en el acceso a un diagnóstico oportuno.

Ante este escenario, los modelos de aprendizaje profundo se han posicionado como una alternativa técnicamente sólida para el apoyo diagnóstico. Brinker et al. [4] demostraron que un

modelo basado en redes neuronales convolucionales superó el rendimiento de 136 de 157 dermatólogos en una tarea de clasificación de imágenes dermatoscópicas de melanoma, lo que evidenció el potencial de la inteligencia artificial como herramienta complementaria al criterio clínico. Sin embargo, las redes neuronales convolucionales (CNNs) presentan una limitación estructural relevante: al operar mediante filtros locales de convolución, tienen dificultades para capturar relaciones de largo alcance dentro de la imagen [7], lo que puede comprometer el diagnóstico en casos donde la estructura global de la lesión —su asimetría, por ejemplo— es determinante.

Los Vision Transformers (ViTs), surgidos como alternativa a las CNN para el análisis de imágenes, abordan precisamente esta limitación. Al procesar la imagen como una secuencia de parches mediante mecanismos de autoatención, los ViTs pueden capturar dependencias globales de manera directa [7]. La integración de ambas arquitecturas en un modelo híbrido CNN–ViT permite aprovechar simultáneamente la eficiencia de las CNNs para extraer características locales y la capacidad de los ViTs para modelar el contexto global, lo que resulta especialmente pertinente en dermatología, donde el diagnóstico requiere evaluar tanto rasgos morfológicos finos como la estructura general de la lesión [5].

A pesar de los avances en el campo internacional, en Colombia los esfuerzos de investigación en cáncer de piel han estado concentrados principalmente en estudios epidemiológicos y guías de detección temprana [15, 16], sin avanzar hacia el desarrollo de herramientas de apoyo diagnóstico que integren inteligencia artificial en prototipos funcionales accesibles. Esta brecha tecnológica, sumada a las desigualdades de acceso a la atención especializada, justifica el desarrollo de soluciones que puedan desplegarse en entornos web y ser utilizadas por personal médico desde cualquier dispositivo con conexión a internet, en línea con

las estrategias de salud digital promovidas por el Ministerio de Salud y Protección Social de Colombia [31].

Bajo estas consideraciones, se plantea la siguiente pregunta de investigación:

¿De qué manera es posible desarrollar un sistema de apoyo al diagnóstico de cáncer de piel basado en modelos híbridos de aprendizaje profundo que permita la clasificación de lesiones cutáneas como benignas o malignas, y cuya integración en una aplicación web facilite su acceso por parte del personal médico?

Para esto se desarrolló un sistema de apoyo al diagnóstico de cáncer de piel basado en un modelo híbrido de aprendizaje profundo integrado en una aplicación web. En este se construyó un conjunto de datos a partir del ISIC Archive 2019, sobre el cual se entrenaron de manera independiente una red neuronal convolucional y un Vision Transformer. Las representaciones extraídas por cada arquitectura se fusionan mediante un módulo Attention Gate que pondera la contribución de cada rama antes de producir la clasificación final. Finalmente, se integra en un prototipo web cliente-servidor que permite al personal médico cargar una imagen dermatoscópica y visualizar la predicción dada por el modelo entrenado.

## 1. Objetivos

### 1.1 Objetivo General

Implementar un prototipo funcional de una aplicación web empleando modelos híbridos de aprendizaje profundo para la clasificación de malignidad de lesiones cutáneas sospechosas de cáncer de piel.

### 1.2 Objetivos Específicos

- Preparar un conjunto de datos de imágenes dermatoscópicas provenientes de repositorios open source.
- Implementar un modelo híbrido CNN / ViT que integre características complementarias de ambas arquitecturas.
- Evaluar el desempeño del modelo híbrido mediante métricas de clasificación médica —precisión, sensibilidad y especificidad— para determinar su robustez y capacidad de generalización.
- Desarrollar un prototipo de aplicación web con arquitectura cliente-servidor que integre el modelo híbrido CNN–ViT en una interfaz funcional posibilitando la interacción entre el usuario y el sistema de clasificación de lesiones cutáneas.
- Validar la integración de los componentes del modelo en el prototipo web mediante pruebas experimentales y de usabilidad con usuarios potenciales.

## 2. Marco Referencial

### 2.1. Cáncer de piel

El cáncer de piel es una enfermedad caracterizada por el crecimiento descontrolado de células anormales en los tejidos cutáneos. Estas células malignas tienen la capacidad de invadir y destruir los tejidos sanos circundantes y, en algunos casos, de desprenderse y propagarse hacia otras partes del cuerpo a través del proceso de metástasis [17].

La mayoría de los cánceres cutáneos se originan en la epidermis (capa más externa de la piel) y se clasifican según el tipo de célula afectada:

- Carcinoma de células basales: se desarrolla en las células basales de la epidermis. Es el tipo más frecuente constituyendo el 80% de todos los cánceres cutáneos en varias poblaciones, de crecimiento lento y con baja probabilidad de metástasis, aunque puede ocasionar daños locales significativos.

- Carcinoma de células escamosas: se origina en las células escamosas de la epidermis. Tiene un mayor potencial de invadir tejidos profundos y de metastatizar en comparación con el carcinoma basocelular.

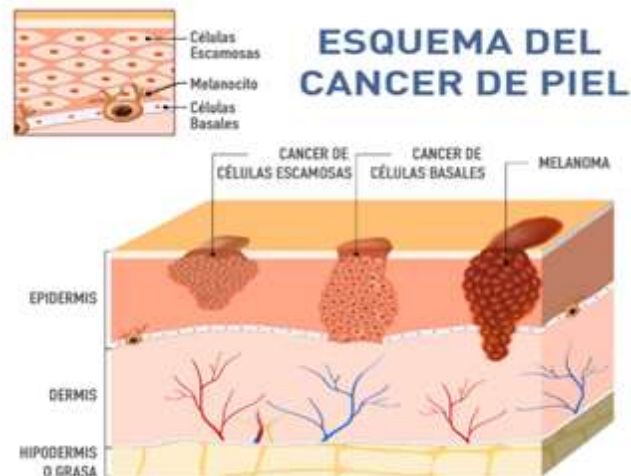
- Melanoma: surge a partir de los melanocitos, células encargadas de la producción de melanina. Es la forma más agresiva y peligrosa de cáncer de piel, debido a su elevada capacidad de diseminación, responsable de cerca del 75 % de las muertes asociadas al cáncer cutáneo, a pesar de representar menos del 5 % de los casos diagnosticados

- Otros tipos menos frecuentes incluyen el carcinoma de células de Merkel, entre otros subtipos cutáneos [18].

En la Figura 1 se puede visualizar algunos de los principales tipos de cancer de piel de forma gráfica.

**Figura 1:**

*Esquema de los principales tipos de cáncer de piel.*



**Nota:** Se representa la localización de los tres tipos más comunes de cáncer cutáneo. Se evidencian el carcinoma de células escamosas en la parte superior de la epidermis, el carcinoma de células basales en las capas más profundas y el melanoma, que puede invadir tejidos más profundos. Adaptado de “Cáncer de piel”, por Policlínico San Miguel, s. f.,

<https://policlinicosanmiguel.com/especialidades-medicas/dermatologia/cancer-de-piel/>

### 2.1.1. Métodos tradicionales de diagnóstico

El diagnóstico del cáncer de piel tradicionalmente se ha basado en una serie de métodos clínicos que permiten la identificación y confirmación de lesiones sospechosas.

En primera instancia se recurre a la inspección visual directa, en la que el médico evalúa las características morfológicas de la lesión mediante herramientas como la regla ABCD (Asimetría, Bordes, Color y Diámetro) o la lista de verificación de siete puntos, ampliamente empleadas en la práctica clínica.

Para mejorar la exactitud diagnóstica, en Estados Unidos y Europa (principalmente Alemania) se introdujo la dermatoscopia o microscopía de epiluminiscencia, un método no invasivo que permite observar estructuras subepidérmicas y patrones de pigmentación invisibles

al ojo desnudo. Numerosos estudios han demostrado que la dermatoscopia incrementa significativamente la sensibilidad y la especificidad del diagnóstico en comparación con el examen clínico tradicional. Por ejemplo, en un meta-análisis que incluyó más de 100 estudios, se encontró que la dermatoscopia aplicada por dermatólogos experimentados ofrecía mayor precisión en la detección de melanoma y carcinomas queratinocitos que el examen visual [20]. De manera similar, estudios más recientes confirman que la dermatoscopia alcanza sensibilidades cercanas al 89 % y especificidades superiores al 90 %, superando al examen clínico en ambos parámetros [21]. No obstante, su efectividad sigue dependiendo de la experiencia del especialista y de la interpretación subjetiva de las imágenes, lo que puede generar variabilidad diagnóstica y limitar la clasificación en contextos con recursos o personal especializado reducido.

Finalmente, la confirmación histopatológica se realiza mediante procedimientos de laboratorio especializados que permiten examinar las características celulares y estructurales de la lesión. Estos métodos posibilitan determinar con precisión el tipo de neoplasia, su grado de invasión y otros aspectos relevantes para el pronóstico y tratamiento [21]. Sin embargo, este procedimiento requiere infraestructura de laboratorio especializada y conlleva tiempos de espera que pueden ser significativos. Por estas razones no puede utilizarse como herramienta de tamizaje primario en la población general, lo que evidencia la necesidad de explorar un mecanismo intermedio, como lo son herramientas automatizadas que apoyen la evaluación objetiva y estandarizada de las lesiones cutáneas.

## **2.2. Aprendizaje profundo aplicado al análisis de imágenes médicas**

Los métodos de aprendizaje profundo han transformado el análisis de imágenes médicas al introducir métodos que permiten la detección automatizada de patrones complejos. En particular, el aprendizaje profundo (deep learning), basado en arquitecturas de redes neuronales profundas,

se ha consolidado como una de las herramientas más eficaces en este campo. A diferencia de los enfoques tradicionales que dependían de la extracción manual de características, el aprendizaje profundo es capaz de aprender representaciones jerárquicas directamente de los datos, lo que permite mejorar la consistencia del análisis y apoyar la toma de decisiones clínicas con menor dependencia de la interpretación subjetiva [22].

En aplicaciones clínicas, el deep learning ha demostrado resultados sobresalientes en radiología, patología digital, oftalmología y dermatología, alcanzando desempeños comparables a los de especialistas humanos en tareas específicas y consolidándose como una herramienta complementaria en el apoyo diagnóstico [4]. Su principal fortaleza radica en la capacidad de procesar grandes volúmenes de datos y extraer patrones ocultos que serían imperceptibles para el ojo humano. Sin embargo, estos modelos también presentan desafíos relacionados con la necesidad de grandes cantidades de datos anotados, la interpretabilidad de los resultados y la validación clínica rigurosa antes de su implementación en la práctica médica [23].

El empleo de estas técnicas en dermatología resulta especialmente prometedor, dado que las lesiones cutáneas pueden documentarse en imágenes dermatoscópicas de alta resolución, disponibles en repositorios de acceso abierto como HAM10000 e ISIC Archive (<https://api.isic-archive.com/collections/66/>). Estas bases de datos han facilitado el desarrollo de modelos robustos para la clasificación de lesiones benignas y malignas, constituyendo un campo activo de investigación y aplicación en salud digital [6].

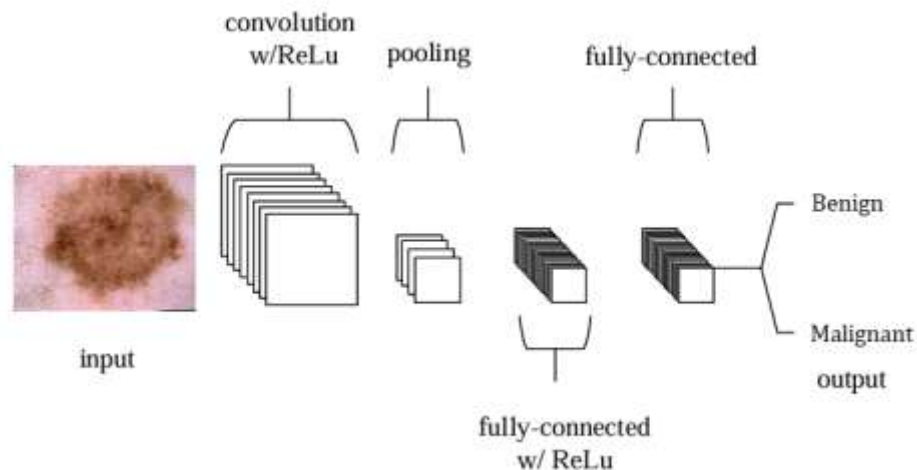
### **2.3. Redes Neuronales Convolucionales (CNNs)**

Las redes neuronales convolucionales (CNNs) son uno de los modelos de aprendizaje profundo más influyentes en el análisis de imágenes médicas, debido a su capacidad para extraer automáticamente características relevantes en diferentes niveles de abstracción [22]. A diferencia

de las redes neuronales tradicionales, las CNNs utilizan capas de convolución que aplican filtros sobre la imagen de entrada para detectar patrones locales, como bordes, texturas o formas, los cuales se combinan progresivamente para reconocer estructuras más complejas. En la figura 2 se ilustra este sistema descrito.

**Figura 2.**

*Ejemplo de estructura básica de una red CNN*



*Nota.* Representa la arquitectura general de una red neuronal convolucional orientada a las lesiones cutáneas, donde se observa el flujo desde la imagen de entrada hasta las capas de convolución, agrupamiento y la red fully-connected que genera la salida de clasificación. Tomada y adaptada de “The CNN Architecture” por Research Gate, Man Xiao. [https://www.researchgate.net/figure/The-CNN-architecture-3\\_fig2\\_382371475](https://www.researchgate.net/figure/The-CNN-architecture-3_fig2_382371475).

En el campo de la dermatología, las CNNs han mostrado un desempeño notable en la clasificación de lesiones cutáneas alcanzando resultados comparables al diagnóstico de especialistas dermatólogos [4]. El éxito de estos modelos radica en su habilidad para procesar imágenes dermatoscópicas de alta resolución, en las que identifican patrones sutiles que pueden pasar desapercibidos en una inspección visual tradicional.

Diversos estudios han explorado arquitecturas avanzadas de CNNs aplicadas al cáncer de piel, integrando además técnicas de segmentación de lesiones para mejorar la precisión del diagnóstico. Por ejemplo, Khan et al. propuso un marco unificado basado en CNNs profundas para la segmentación y clasificación de lesiones, demostrando que estos métodos pueden alcanzar altos niveles de sensibilidad y especificidad en la identificación de melanomas [8].

No obstante, las CNNs presentan ciertas limitaciones conceptuales inherentes a su arquitectura. Su diseño se centra predominantemente en la captura de patrones locales mediante operaciones de convolución, lo que puede restringir su capacidad para modelar dependencias globales y relaciones contextuales de largo alcance dentro de la imagen [23].

#### **2.4. Vision Transformers (ViTs)**

Los Vision Transformers (ViTs) representan una evolución en el campo del aprendizaje profundo para el análisis de imágenes. A diferencia de las redes convolucionales, que procesan las imágenes a través de filtros locales, los ViTs dividen la imagen en pequeñas regiones denominadas patches, las cuales son tratadas como secuencias de entrada que se procesan mediante mecanismos de autoatención [7]. Este enfoque permite a los modelos capturar relaciones de largo alcance entre diferentes regiones de la imagen, ofreciendo una visión global del patrón que resulta especialmente útil en contextos donde la información relevante se distribuye de forma no localizada.

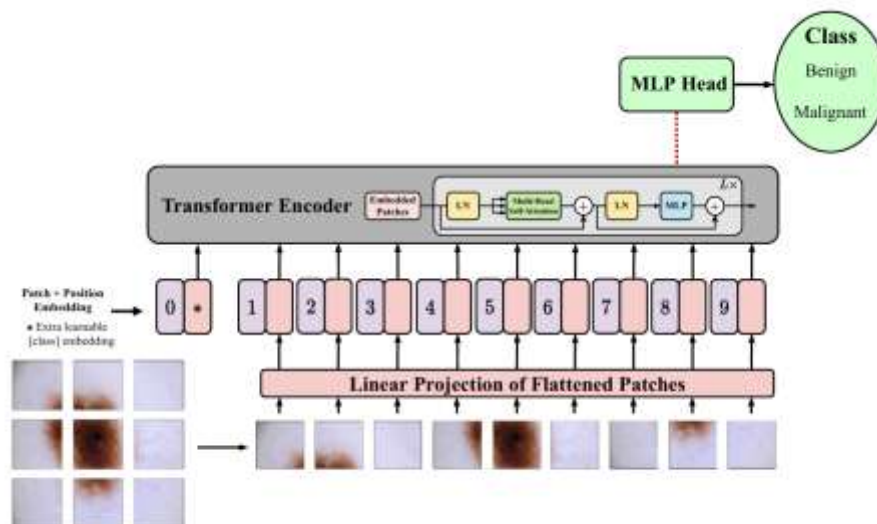
El núcleo de los ViTs reside en el mecanismo de autoatención (self-attention), el cual permite al modelo ponderar la importancia de cada parche en relación con todos los demás, capturando dependencias de largo alcance y contextos globales dentro de la imagen. Este enfoque contrasta con la naturaleza local y progresiva de las CNNs, ofreciendo una capacidad superior para integrar información distribuida y modelar interacciones complejas entre regiones distantes.

En dermatología, los ViTs han mostrado resultados prometedores al clasificar melanomas y otras lesiones cutáneas con precisiones comparables o superiores a modelos convolucionales, especialmente cuando se entrenan o afinan con datasets especializados como los ISIC [6,5].

En la Figura 3 se presenta la arquitectura de los Vision Transformers junto con la estructura detallada del bloque codificador, donde se observa cómo los parches son proyectados y posteriormente procesados.

### Figura 3.

*Arquitectura del Visión Transformer y estructura detallada del bloque codificador*



**Nota.** Se evidencia el funcionamiento interno de un modelo *ViT* orientado a las lesiones cutáneas donde la imagen se divide en parches proyectados linealmente y procesados por capas de atención del transformador para obtener la clasificación. Adaptado de “Vision transformers in medical computer vision—A survey”, por I. Chaves, M. Talo, & P. F. Jaeger, 2023, arXiv, <https://arxiv.org/abs/2301.03505>

### 2.5. Modelos híbridos CNN + ViT

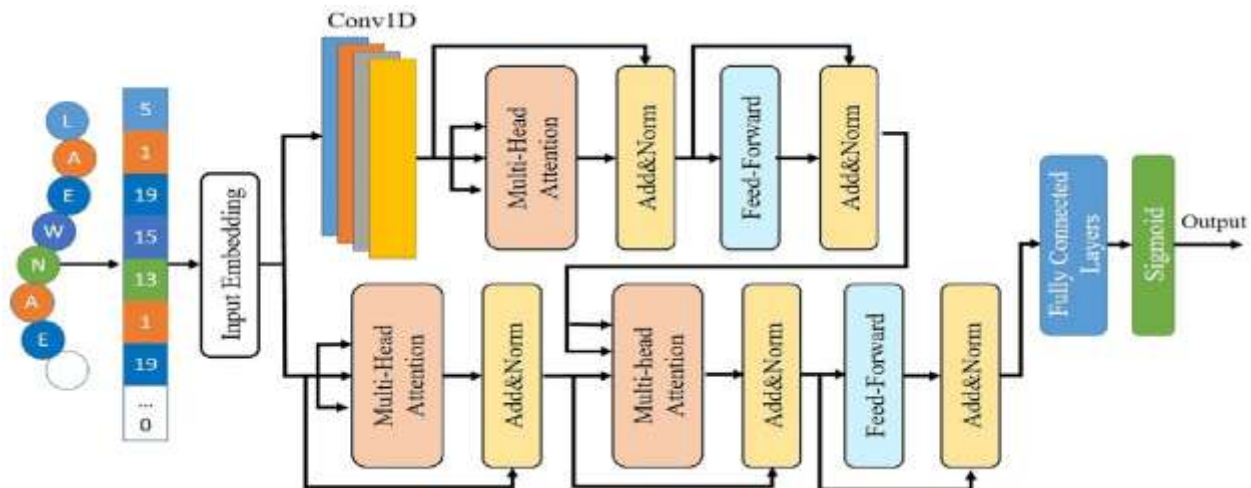
Los modelos híbridos que combinan Redes Neuronales Convolucionales (CNNs) y Vision Transformers (ViTs) integran de manera sinérgica sus ventajas complementarias: las CNNs

aportan una capacidad probada para extraer características locales y patrones espaciales jerárquicos —como texturas, bordes y estructuras morfológicas finas—, mientras que los ViTs contribuyen con su mecanismo de autoatención para capturar dependencias globales y relaciones contextuales de largo alcance dentro de la imagen. Esta combinación permite que el modelo resultante no solo identifique detalles morfológicos sutiles, sino que también analice la estructura general y la disposición espacial de las lesiones, mejorando así la precisión y robustez en tareas de clasificación de imágenes dermatoscópicas [5,7].

En la Figura 4 se muestra el esquema general del modelo híbrido CNN-Transformer para predicción, donde se observa cómo ambas ramas procesan la información y posteriormente la fusionan para generar la salida final.

**Figura 4.**

*Esquema del modelo híbrido Transformer-CNN para predicción.*



**Nota.** Representación de un modelo híbrido que integra capas convolucionales y módulos de atención del ViT. Adaptado de “Hybrid transformer-CNN model for accurate prediction of peptide hemolytic potential”, por S. Almotairi, E. Badr, I. Abdelbaky, M. Elhakeem, & M. Abdul Salam, 2024, Scientific Reports, 14, 14263, <https://doi.org/10.1038/s41598-024-63446-5>

En el contexto de la dermatología, la identificación de lesiones cutáneas malignas depende tanto de detalles locales específicos (por ejemplo, irregularidades en el borde o variaciones de pigmentación) como de la estructura global de la lesión. Los modelos híbridos permiten aprovechar simultáneamente ambas dimensiones, lo que se traduce en un rendimiento superior respecto a arquitecturas individuales [4,8]. Alcanzando resultados más consistentes en datasets abiertos como el ISIC Archive, superando tanto a CNNs profundas como a ViTs entrenados de manera aislada [6].

## **2.6. Aprendizaje profundo aplicado a la clasificación de lesiones cutáneas**

El uso de modelos de aprendizaje profundo para el diagnóstico asistido del cáncer de piel cuenta con una trayectoria de investigación consolidada a nivel internacional. El trabajo de Brinker entre otros, marcó un punto de inflexión al demostrar que un modelo CNN superó a la mayoría de los especialistas participantes en una tarea de clasificación de imágenes dermatoscópicas de melanoma [4]. Este resultado evidenció el potencial de la inteligencia artificial como herramienta complementaria al criterio clínico y motivó una ola de investigaciones orientadas a explorar, refinar y contextualizar este tipo de sistemas. En la misma línea, Khan et al. [8] desarrollaron un marco unificado para la segmentación y clasificación de lesiones cutáneas en el dataset HAM1000, demostrando que la integración de ambas tareas en un mismo pipeline mejora la precisión diagnóstica.

En cuanto a los Vision Transformers aplicados a dermatología, estudios posteriores al trabajo seminal de Dosovitskiy los adaptaron a imágenes dermatoscópicas con resultados prometedores, especialmente en casos clínicamente ambiguos donde la estructura global de la lesión es determinante [6, 7]. Más recientemente, Akinrinade y Du confirmaron que las arquitecturas híbridas CNN-ViT superan en desempeño a los enfoques individuales en el ISIC

Archive [5], consolidando una tendencia hacia modelos más robustos y complementarios en el ámbito clínico.

Posteriormente, estudios como los de Haggemüller aportaron una perspectiva crítica al analizar el desempeño de CNNs en clasificación de cáncer de piel frente a dermatólogos, concluyendo que, si bien los modelos pueden alcanzar niveles de precisión comparables en contextos controlados, su implementación clínica a gran escala enfrenta desafíos relacionados con la generalización, la interpretabilidad y la integración en los flujos de trabajo médicos establecidos [22]. La literatura de Tschandl entre otros, complementaron este panorama al explorar esquemas de colaboración entre dermatólogos y sistemas automatizados, demostrando que la sinergia entre el criterio clínico especializado y los algoritmos de aprendizaje profundo enriquece el proceso diagnóstico más allá de lo que cada componente logra de manera aislada [23].

## **2.7. Diagnóstico asistido por IA en Colombia**

En Colombia y América Latina los esfuerzos de investigación sobre cáncer de piel han estado concentrados principalmente en estudios epidemiológicos y guías de detección temprana. La incidencia del cáncer de piel en Colombia ha mostrado un aumento sostenido en la última década [14]. Bonilla entre otros, analizaron su comportamiento entre 2003 y 2007, reportando tasas crecientes en diversas regiones del país y resaltando la necesidad de programas de prevención y diagnóstico oportuno [15]. El Instituto Nacional de Cancerología desarrolló un manual para la detección temprana del cáncer de piel orientado a estandarizar la atención y capacitar al personal de salud [16]. Sin embargo, los esfuerzos de investigación aplicada que integren modelos de inteligencia artificial en herramientas funcionales de apoyo diagnóstico son prácticamente inexistentes en el contexto nacional y escasos en la región. La gran mayoría de los trabajos

disponibles son de carácter epidemiológico, sin avanzar hacia el desarrollo de prototipos validados localmente.

### **3. Método propuesto**

En este capítulo se describe detalladamente la metodología empleada para construir el sistema de apoyo al diagnóstico de cáncer de piel, con el objetivo de desarrollar una herramienta de apoyo accesible y funcional para los dermatólogos.

Este capítulo se organiza en cinco secciones, donde la primera sección –3.1– describe la preparación e integración de los conjuntos de datos tomado del dataset ISIC. La segunda sección –3.2– describe las métricas de evaluación empleadas en los modelos. En la tercera y cuarta sección –3.3 y 3.4– se presenta la arquitectura, el entrenamiento y las métricas de evaluación que rigen ambos modelos individuales, tanto el modelo convolucional *CNN* como el modelo Vision Transformer *ViT* respectivamente; la quinta sección –3.5– desarrolla y detalla la arquitectura, el entrenamiento y la técnica de fusión de características del modelo híbrido *CNN + ViT* apoyado de los pesos resultantes de los mejores modelos individuales experimentados. En la sexta sección –3.6– se describe la construcción del sistema web con arquitectura cliente-servidor junto a la integración del modelo híbrido el cual pone las predicciones de este en funcionamiento para la disposición y evaluación del usuario final. Por último, en la sección –3.7– se hace una discusión de los resultados dados en cada uno de las anteriores secciones.

#### **3.1. Preparación de datos**

##### ***3.1.1. Dataset ISIC 2019***

El punto de partida de un sistema de diagnóstico con aprendizaje profundo es contar con una buena cantidad de datos bien etiquetados y representativos. En el caso de las lesiones cutáneas,

esto se obtiene gracias a datasets públicos de descarga libre para uso académico como los que ofrece el desafío anual de la International Skin Imaging Collaboration – ISIC –, cuya misión es promover el desarrollo y la evaluación comparativa de herramientas automáticas para el análisis dermatoscópico.

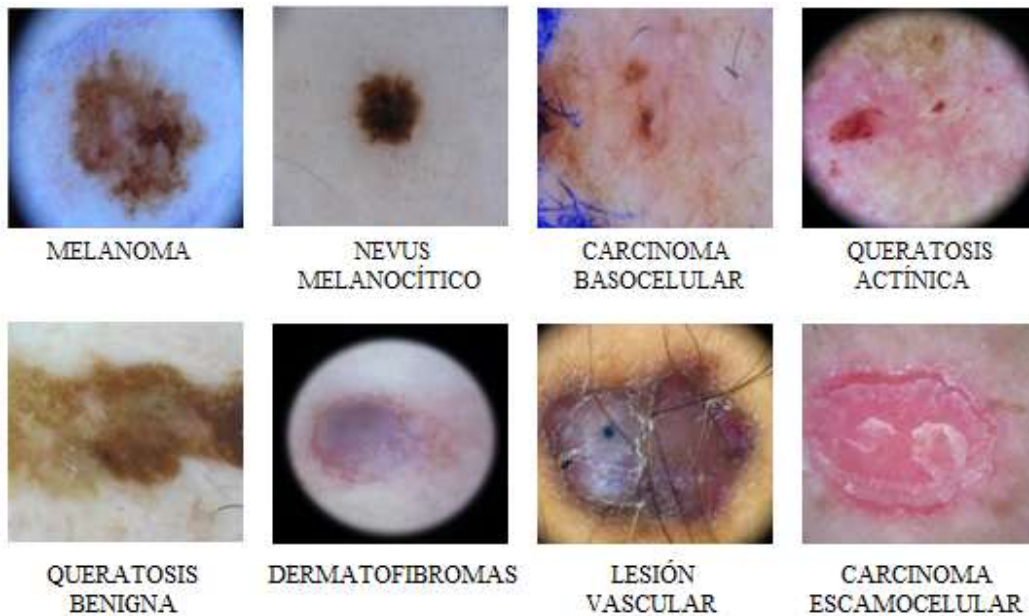
Desde su primera edición en 2016, el desafío ISIC ha consolidado la mayor colección pública de imágenes dermatoscópicas del mundo, convirtiéndose en el corpus de referencia estándar para la comunidad de aprendizaje profundo en esta área [23].

La edición de 2019 del desafío, denominada ISIC 2019, provee un conjunto de 25.331 imágenes dermatoscópicas distribuidas en ocho categorías diagnósticas: melanoma (MEL), nevus melanocítico (NV), carcinoma basocelular (BCC), queratosis benigna (BKL), queratosis actínica (AK), carcinoma escamocelular (SCC), lesiones vasculares (VASC) y dermatofibroma (DF). Las imágenes fueron capturadas en tres instituciones dermatológicas de referencia —el Hospital Universitario de Zurich, el Medical University of Vienna y el Memorial Sloan Kettering Cancer Center— utilizando cámaras dermatoscópicas calibradas y etiquetadas por dermatólogos con base en diagnóstico histopatológico confirmado o en criterios de consenso clínico revisados.

En la Figura 5 se presentan ejemplos representativos de las distintas categorías incluidas en el conjunto de datos ISIC 2019, ilustrando la variabilidad morfológica y visual existente entre las diferentes clases diagnósticas.

**Figura 5.**

*Ejemplos de las diversas categorías en el dataset ISIC 2019.*



*Nota.* Adaptado de “Application of Semi-Supervised Learning in Image Classification: Research on Fusion of Labeled and Unlabeled Data”, por Sai Li, Peng Kou & Miao Ma. [https://www.researchgate.net/figure/SIC-2019-Dataset-Analysis-of-the-Distribution-of-Sample-Samples-of-Variou-Categories\\_fig4\\_378350495](https://www.researchgate.net/figure/SIC-2019-Dataset-Analysis-of-the-Distribution-of-Sample-Samples-of-Variou-Categories_fig4_378350495)

Desde el punto de vista clínico, el objetivo prioritario de un sistema de apoyo al diagnóstico en atención primaria no es necesariamente determinar el tipo exacto de lesión sino identificar con alta sensibilidad si la lesión pertenece a una categoría que requiere atención urgente o especializada. Con esta perspectiva, las ocho categorías diagnósticas se agruparon en dos clases binarias: maligna y benigna. En el grupo maligno se incluyeron el melanoma (MEL), el carcinoma basocelular (BCC), la queratosis actínica (AK) y el carcinoma escamocelular (SCC). El melanoma es el tipo de cáncer de piel con mayor mortalidad dado su alto potencial metastásico; el carcinoma basocelular y el escamocelular son los más frecuentes y, aunque raramente mortales, requieren

tratamiento quirúrgico oportuno; la queratosis actínica es una lesión premaligna con potencial documentado de progresión a carcinoma escamocelular. En el grupo benigno se incluyeron el nevus melanocítico (NV), la queratosis benigna (BKL), el dermatofibroma (DF) y las lesiones vasculares (VASC), todas con pronóstico generalmente favorable y que en la mayoría de los casos no requieren intervención urgente.

Para que se cumpla esta dicotomía, esto se implementa mediante la variable label, definida por el diccionario LABEL\_MAP donde las lesiones benignas se identifican con un valor “0”, mientras que las lesiones malignas se representan con el valor “1”; resultando una diferenciación plasmada que se resume en la Tabla 1.

**Tabla 1. Distribución de imágenes por categoría binaria en el corpus ISIC 2019**

Categoría	Clases diagnósticas	N imágenes	% del corpus
Benigno (0)	NV, BKL, DF, VASC	15.991	63,1%
Maligno (1)	MEL, BCC, AK, SCC	9.340	36,9%

*Nota.* Elaboración a partir del archivo ISIC\_2019\_Training\_GroundTruth.csv, donde MEL = melanoma; BCC = carcinoma basocelular; AK = queratosis actínica; SCC = carcinoma escamocelular; NV = nevus melanocítico; BKL = queratosis benigna; DF = dermatofibroma; VASC = lesión vascular.

### 3.1.2. Construcción, integración y alteración de dataset para entrenamiento

El proceso inicia obteniendo la etiqueta de las ocho clases diagnósticas por imagen y excluyendo las instancias (UNK) que no pertenecen a ninguna de las ocho categorías principales y no son aptas para el entrenamiento supervisado. Posteriormente, se fusionan los metadatos, se

resuelven las rutas absolutas y se aplica el mapeo binario, generando un Dataframe que contiene ruta de imagen, etiqueta diagnóstica original, etiqueta binaria y una clave de estratificación, que actúa como única fuente de verdad –ground truth–. Posteriormente, el dataframe obtenido se divide en entrenamiento, validación y prueba mediante partición estratificada “70 / 15 / 15” respectivamente, garantizando la preservación de la distribución de clases y orígenes y la reproducibilidad del experimento. Las divisiones resultantes con valores absolutos de imágenes por clase se exponen en la Tabla 2.

**Tabla 2. Particiones del corpus ISIC**

<b>Partición</b>	<b># Imágenes Totales</b>	<b>Benignas</b>	<b>Malignas</b>
Entrenamiento (70%)	17.731	11.193	6.538
Validación (15%)	3.800	2.399	1.401
Prueba (15%)	3.800	2.399	1.401

*Nota.* División estratificada por clase manteniendo proporción benigna/maligna del dataset original.

Luego, para el conjunto de entrenamiento se aplica una secuencia de transformaciones que combina aumento de datos aleatorio – data augmentation – y normalización, donde las imágenes se alteran con rotaciones horizontal y vertical, rotación  $\pm 20^\circ$  y variación de brillo/ contraste/ saturación/ tono. El aumento de datos es una práctica estándar en clasificación de lesiones cutáneas y ha demostrado reducir el sobreajuste en corpus de tamaño moderado [34]. Una vez transformadas, estas se redimensionan a 224×224 píxeles resolución estándar de los backbones preentrenados en ImageNet que se utilizan como extractores de características. La interfaz con

PyTorch se implementa en la clase `SkinDataset`, que carga cada imagen, aplica las transformaciones y devuelve el par (imagen, etiqueta binaria).

Finalmente, para contrarrestar el desbalance de clase descrito en la Tabla 1, se calculan pesos por clase mediante la fórmula

$$W_c = \frac{n_{total}}{n_{clases} * n_c},$$

donde  $n_{total}$  es el número total de imágenes de entrenamiento,  $n_{clases}$  es 2 y  $n_c$  es el número de imágenes de la clase  $c$  en el conjunto de entrenamiento. El peso resultante es mayor para la clase maligna, la cual está menos representada, y menor para la clase benigna, de modo que el gradiente de los ejemplos malignos contribuye de una manera más proporcional al ajuste de los parámetros a la hora del entrenamiento.

### 3.2. Métricas de evaluación

Para evaluar el desempeño de los tres modelos — CNN, ViT e Híbrido —se utilizó un conjunto común de métricas de clasificación binaria. La elección de estas métricas responde tanto a la naturaleza del problema —detección de lesiones malignas— como a los requerimientos clínicos del sistema, donde el costo de un falso negativo es considerablemente mayor que el de un falso positivo.

La exactitud (accuracy) mide la proporción de predicciones correctas sobre el total de casos. Sin embargo, esta resulta insuficiente cuando las clases están desbalanceadas, como ocurre en este dataset, donde las imágenes benignas superan el 60% del total.

La sensibilidad, también llamada recall de la clase maligna o tasa de verdaderos positivos, mide qué proporción de los casos malignos reales fueron correctamente identificados por el

modelo. Es la métrica de mayor relevancia clínica, ya que implica la capacidad para detectar lesiones que requieren atención urgente y que no pasen desapercibidas.

La especificidad, o recall de la clase benigna, mide qué proporción de los casos benignos reales fueron clasificados correctamente. Un valor alto de especificidad reduce las remisiones innecesarias de pacientes con lesiones benignas, lo cual tiene un impacto directo en la eficiencia del sistema de salud.

La precisión indica qué proporción de los casos predichos como malignos lo son realmente. Una precisión baja implica un alto número de falsos positivos.

Se agrega a estas métricas el F1-score, que es la media armónica entre precisión y sensibilidad, ofreciendo un balance entre ambas cuando existe desbalance de clase.

Finalmente, el AUC-ROC (área bajo la curva ROC) cuantifica la capacidad discriminativa global del modelo a lo largo de todos los posibles umbrales de decisión, siendo independiente del umbral elegido y por tanto útil para comparar modelos de forma objetiva. Un AUC de 1,0 indica discriminación perfecta, por lo que el valor hallado tiene que ser lo más próximo posible a esta cifra.

En la Tabla 3 se ve la definición práctica de cada una de estas métricas, junto a un pequeño resumen de la relevancia clínica que tiene cada una.

**Tabla 3. Métricas de evaluación clínica empleadas.**

Métrica	Definición	Relevancia clínica
Exactitud (Accuracy)	$\frac{VP + VN}{Total}$	Proporción global de predicciones correctas.
Sensibilidad (Recall)	$\frac{VP}{VP + FN}$	Detectar lesiones malignas.
Especificidad	$\frac{VN}{VN + FP}$	Evitar alarmas en casos benignos.
Precisión	$\frac{VP}{VP + FP}$	Confiabilidad de las predicciones positivas.
F1-score	$\frac{2 * (Precisión) (Sensibilidad)}{Precisión + Sensibilidad}$	Balance entre precisión y sensibilidad.
AUC-ROC	Área bajo la curva ROC	Desempeño global a todos los umbrales.

**Nota.** Donde,  $VP$  = verdaderos positivos;  $VN$  = verdaderos negativos;  $FP$  = falsos positivos;  $FN$  = falsos negativos.

### 3.2.1. Selección del umbral óptimo

Además de la evaluación con umbral  $\tau = 0,5$ ; para determinar el umbral óptimo de un modelo se hace el uso del índice de Youden. Este es un parámetro estadístico que evalúa la efectividad global de una prueba diagnóstica para determinar el punto de corte óptimo.

$$J = Sensibilidad + Especificidad - 1$$

Este índice maximiza la suma de sensibilidad y especificidad de forma simultánea, proporcionando un punto de la curva ROC que equilibra ambas métricas. El umbral óptimo se identificó como el valor de  $\tau$  que maximiza  $J$  sobre el conjunto de validación.

### ***3.2.2. Intervalos de confianza bootstrap***

Para el modelo híbrido se calcularon intervalos de confianza al 95% para todas las métricas clave mediante bootstrap no paramétrico con 1.000 remuestreos sobre el conjunto de prueba ( $n = 3.800$ ), siguiendo la metodología propuesta por [35]. En cada iteración se extrajo una muestra aleatoria con reemplazo del mismo tamaño que el conjunto de prueba y se calculó la métrica de interés. Los percentiles 2,5 y 97,5 de la distribución de los 1.000 valores estimados definen el intervalo de confianza al 95%. Este procedimiento cuantifica la variabilidad de las métricas reportadas y permite establecer comparaciones estadísticas.

### ***3.2.3. Calibración probabilística***

La calibración de un clasificador describe el grado de correspondencia entre la probabilidad predicha y la frecuencia observada de la clase positiva. Un modelo bien calibrado se ve representado como aquel en el que, entre todos los ejemplos a los que asigna una probabilidad de malignidad de  $p$ , aproximadamente una fracción  $p$  efectivamente corresponda a casos malignos. En nuestro caso, la calibración tiene relevancia clínica directa ya que la probabilidad continua de malignidad es la que orienta la urgencia de derivación.

Dicho esto, se implementó la calibración en el modelo híbrido mediante el Expected Calibration Error (ECE), calculado como el promedio ponderado de la discrepancia entre confianza media y exactitud observada en diez intervalos de probabilidad equiespaciados [36].

Según la literatura, Un Expected Calibration Error inferior a 0,05 se considera un modelo excelentemente calibrado, entre 0,05 y 0,10 un modelo con buena calibración, mientras que un modelo superior a 0,10 requiere corrección.

### **3.3. Modelo CNN**

La primera rama del modelo híbrido corresponde a una red convolucional EfficientNet-B0 preentrenada en ImageNet. Esta sección describe la arquitectura, el proceso de búsqueda de hiperparámetros, la configuración final seleccionada y los algunos resultados sobre el conjunto de prueba.

#### ***3.3.1. Selección de arquitectura e hiperparámetros***

La literatura reciente, incluyendo benchmarks de 2024 y 2025 que comparan hasta 38 arquitecturas CNN diferentes sobre HAM10000 e ISIC, muestra que no existe una única arquitectura dominante para la clasificación de lesiones cutáneas. Los resultados de EfficientNet, ResNet, DenseNet y otras variantes caen dentro de un rango estrecho de rendimiento, típicamente entre el 85% y el 94% de accuracy, por lo que la elección específica depende del dataset, el esquema de transfer learning y los hiperparámetros del experimento. Por esta razón, la contribución del presente proyecto no busca identificar la mejor CNN individual existente, sin embargo se hicieron experimentos para hallar la mejor combinación posible tanto de arquitectura como de hiperparámetros para que el modelo híbrido posteriormente construido se alimente de buenos modelos individuales según las métricas de evaluación.

La red convolucional experimentada se implementó mediante la biblioteca timm [37], donde la salida del modelo corresponde al vector de 1280 dimensiones producido por la agrupación global promedio. Sobre este vector de características se añade un encabezado de clasificación

personalizado compuesto por una capa lineal con activación GELU, un dropout configurable y alterable como hiperparámetro y una capa lineal final a 2 salidas correspondientes a las dos clases del problema de clasificación binaria (benigno/maligno).

Una vez construida la arquitectura base se experimentó iterativamente sobre familias de redes junto a diversos hiperparámetros con el fin de optimizar el modelo convolucional — registrados en el archivo `experiments_cnn.csv` del cuaderno de modelado—, algunos de los resultados para definir el modelo final individual se sintetizan en la Tabla 4.

**Tabla 4. Experimentación Modelo CNN**

Exp	Modelo base	Optimizador	LR	Dropout	Función de pérdida	Sensibilidad	F1	AUC - ROC
1	EffNet	AdamW	$1e^{-4}$	0,5	BCE	0.8715	0.8568	0.9536
2	EffNet	AdamW	$1e^{-4}$	0,3	BCE	0.8715	0.8718	0.9593
3 ✓	EffNet	AdamW	$5e^{-4}$	0,8	BCE.	0.8808	0.8746	0.9628
4	EffNet	AdamW	$3e^{-3}$	0,8	BCE	0.8315	0.8204	0.9398
5	EffNet	AdamW	$5e^{-4}$	0,8	FOCAL	0.8587	0.8633	0.9630
6	EffNet	SGD	$5e^{-4}$	0,8	BCE	0.7823	0.7580	0.8917
7	EffNet	SGD	0.01	0,8	BCE	0.8480	0.8375	0.9468
8	EffNet	Adam	$5e^{-4}$	0,8	BCE	0.8737	0.8758	0.9634

9	ResNet	Adam	$1e^{-4}$	0,8	BCE	0.8030	0.8340	0.9442
10	ResNet	Adam	$3e^{-3}$	0,3	BCE	0.7366	0.7727	0.9113
11	ResNet	AdamW	$3e^{-3}$	0,5	FOCAL	0.6916	0.7576	0.9093
12	EffNet	Adam	$5e^{-4}$	0,8	BCE.	0.8708	0.8752	0.9641

---

**Nota.** ✓ Configuración final seleccionada para proveer pesos al modelo híbrido. Experimentos totales hechos en el NB2-Modelo\_CNN y guardados en experiments\_cnn.csv, donde LR = Learning Rate. Todos los experimentos usaron BCEWithLogitsLoss, CosineAnnealingLR

### 3.3.2. Entrenamiento

Con base en los resultados de la Tabla 4, se seleccionó el Experimento 3 como configuración final tomando en cuenta el valor de sensibilidad, crítico en el tema abordado. El experimento escogido tiene la sensibilidad más alta empatada con el experimento 12 (0.8808), sin embargo, posee mejor especificidad y F1 que dicho experimento. Por otro lado, el experimento 3 no es el mejor en cuanto a AUC-ROC, siendo este el experimento 12, no obstante, la diferencia entre 0.9628 y 0.9641 es de 0.0013, estadísticamente insignificante. En términos clínicos, el experimento 3 detecta correctamente un aproximado de 88 de cada 100 malignos; mientras que, el experimento 12 detecta aproximadamente 87 de la misma cifra base. Esa diferencia de 1 maligno adicional detectado por cada 100 es un argumento más sólido que 0.001 de valor AUC - ROC.

El entrenamiento completo se realizó con el optimizador AdamW, una tasa de aprendizaje  $5 \times 10^{-4}$ , decaimiento de pesos  $1 \times 10^{-4}$ , tamaño de lote de 64, Dropout de 0,8; Scheduler CosineAnnealingLR y función de pérdida BCEWithLogitsLoss, que combina en una sola

operación numéricamente estable la función sigmoide y la entropía cruzada binaria, evitando problemas de saturación del gradiente [38], todo durante 25 epochs. Este proceso se ejecutó con precisión mixta (AMP) en GPU T4 x2, con un tiempo total de entrenamiento de alrededor de 97 minutos. Toda la configuración descrita se expone en la Tabla 5.

**Tabla 5. Configuración final Modelo CNN**

Hiperparámetro	Valor
Backbone	EfficientNet-B0
Capa oculta del encabezado	Linear (1280→512) → GELU → Dropout (0,5) → Linear (512→1)
Número de salidas	1 sigmoide → 2 clases – benignant / malignant –
Optimizador	AdamW
Tasa de aprendizaje	$5 \times 10^{-5}$
Weight Decay	$1 \times 10^{-4}$
Epochs	25
Función de pérdida	Entropía cruzada con pesos por clase (BCEWithLogitsLoss)
Planificador LR	CosineAnnealingLR
Acelerador	GPU T4 x2

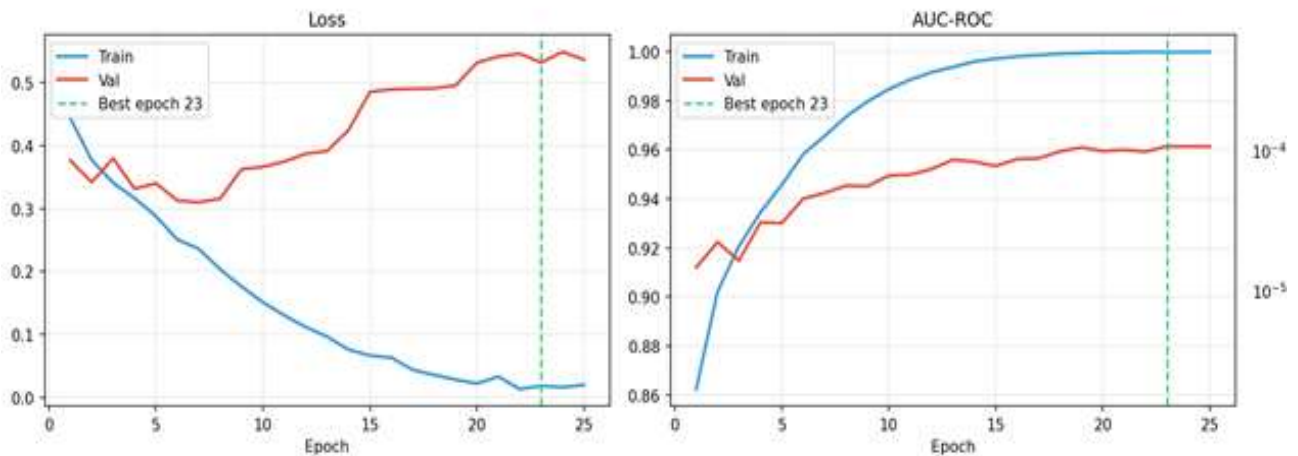
**Nota.** Elaboración a partir de la configuración del notebook 2-modelo-cnn.ipynb.

### 3.3.3. Resultados

La Figura 6 presenta las curvas de entrenamiento del experimento seleccionado, cuyos pesos fueron guardados como artefacto de análisis visual.

#### Figura 6.

Curvas de entrenamiento — CNN EfficientNet-B0



**Nota.** Izquierda: pérdida (BCE) de entrenamiento y validación por epoch. Derecha: AUC-ROC de entrenamiento y validación.. La línea punteada verde marca la epoch de mejor AUC de validación – epoch 23 –.

En la curva de pérdida (panel izquierdo) se aprecia el patrón clásico del ajuste fino desde preentrenamiento: la pérdida de entrenamiento desciende de forma sostenida (de 0,443 en la epoch 1 a 0,020 al final), mientras la pérdida de validación, después de descender hasta la epoch 5, comienza a incrementarse progresivamente a partir de la epoch 15, alcanzando 0,537 al final. Esto, aunque puede indicar un sobreajuste en los últimos epochs, no provoca pérdida de valor AUC - ROC el cual continúa de forma ascendente a pesar del incremento en la pérdida de validación, reflejando que el modelo sigue mejorando su capacidad discriminativa global.

En la curva AUC-ROC (panel derecho) el AUC de entrenamiento asciende rápidamente hasta valores muy cercanos a 1,00, mientras el AUC de validación crece de 0,912 en la epoch 1 a 0,949 en la epoch 10, y 0,9628 en la epoch 23.

**Tabla 6. Reporte de clasificación modelo CNN**

Clase	Precisión	Sensibilidad	F1-score	Soporte (Imágenes)
Benigno	0,9250	0,9304	0,9277	2399
Maligno	0,8796	0,8808	0,8801	1401
Exactitud global	—	0,9084	—	3800
AUC - ROC	—	0,9628	—	—

**Nota.** Métricas calculadas con sklearn.metrics. El umbral óptimo de Youden fue  $\tau^* = 0,405$ , con el cual se maximiza la suma sensibilidad + especificidad.

La Tabla 6 presenta el reporte de clasificación del modelo EfficientNet utilizado, donde se alcanzó una exactitud global de 0,9084; clasificando correctamente 3.446 de las 3.800 imágenes totales. Para la clase benigna, la precisión (0,9250) y la sensibilidad (0,9304) son simétricamente altas, lo cual, acompañado con un F1-score de 0,9277 nos indica que el modelo identifica con solidez las lesiones que no requieren derivación urgente.

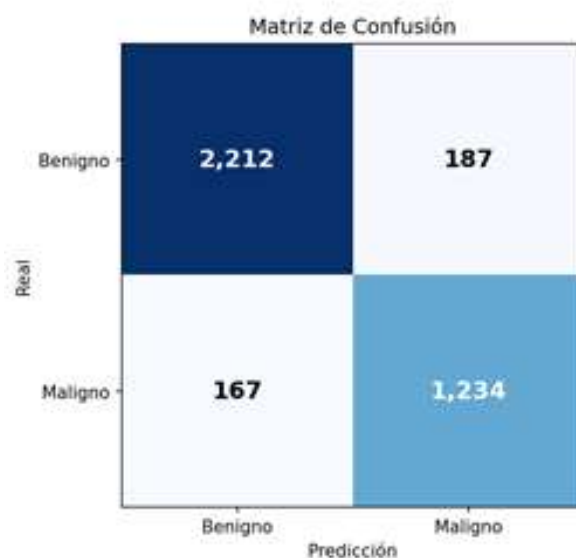
Como se explicó anteriormente, la métrica de mayor relevancia clínica es la sensibilidad de la clase maligna (0,8808); la cual indica que el modelo detectó correctamente el 88,1% de los 1.401 casos malignos del conjunto de prueba. La precisión de la clase maligna (0,8796) muestra

que el 87,9% de las predicciones positivas del modelo corresponden efectivamente a casos malignos.

Entendidas las métricas, se reporta en la Figura 7 la matriz de confusión con todas las predicciones dadas por el modelo.

**Figura 7.**

*Matriz de confusión - Modelo CNN*



**Nota.** Matriz de confusión con umbral  $\tau = 0,5$ ; valores absolutos sobre el conjunto de prueba.

Esta matriz de confusión muestra los resultados absolutos sobre las 3.800 imágenes: 2.212 benignos correctamente clasificados (VN), 187 benignos predichos erróneamente como malignos (FP), 167 malignos clasificados como benignos (FN) y 1.234 malignos correctamente identificados (VP). Los 167 falsos negativos representan el error de mayor consecuencia clínica, equivalente al 11,9% de los casos malignos del conjunto de prueba, por lo que es la métrica que se busca mejorar.

### 3.4. Modelo *VIT*

La segunda rama corresponde a una arquitectura de tipo Vision Transformer. A diferencia de la CNN, estos modelos procesan la imagen mediante mecanismos de autoatención global sin el sesgo inductivo de localidad espacial, capturando dependencias entre regiones distantes de la imagen desde la primera capa. La exploración experimental cubrió dos familias de esta arquitectura: ViT-B/16 [39] y DeiT-B/16 [40], con el objetivo de identificar la configuración con mejor balance clínico sobre el corpus ISIC 2019.

#### 3.4.1. Selección de arquitectura e hiperparámetros

Tanto ViT-B/16 como DeiT-B/16 comparten la misma estructura de procesamiento: dividen la imagen de  $224 \times 224$  en parches no solapados de  $16 \times 16$  píxeles —196 parches en total—, los proyectan linealmente a vectores de 768 dimensiones y procesan la secuencia de 197 tokens mediante 12 bloques Transformer con 12 cabezas de autoatención; Produciendo un vector sobre el que se añade el mismo encabezado que en la CNN: Dropout + Linear(768, 1).

Ambas arquitecturas fueron entrenadas en ImageNet con supervisión estándar, sin embargo, mientras ViT-B/16 usa ImageNet-21k, DeiT-B/16 hace su pre-entrenamiento en ImageNet-1k.

Una vez conocidas las arquitecturas, se realizaron experimentos iterativos de exploración variando hiperparámetros como el optimizador (AdamW, Adam y SGD), la tasa de aprendizaje —learning rate—, la función de pérdida (BCE y Focal) y el número de epochs —registrado en `experiments_vit.csv`—. La Tabla 7 presenta los experimentos hechos y da a entender cuál fue el seleccionado según algunas de las métricas propuestas anteriormente.

**Tabla 7. Experimentación Modelo ViT**

Exp	Modelo base	Optimizador	LR	Dropout	Función de pérdida	Sensibilidad	F1	AUC - ROC
1	ViT - B/16	AdamW	$5 \times 10^{-4}$	0,5	BCE	0,8808	0,8897	0,9677
2	ViT - B/16	AdamW	$5 \times 10^{-3}$	0,5	BCE	0,8294	0,6756	0,8172
3	ViT - B/16	AdamW	$1 \times 10^{-3}$	0,8	BCE	0,8323	0,8229	0,9319
4	ViT - B/16	AdamW	$1 \times 10^{-4}$	0,5	FOCAL	0,8787	0,8802	0,9676
5	ViT - B/16	SGD	$1 \times 10^{-3}$	0,5	BCE	0,7009	0,7468	0,8961
6	ViT - B/16	SGD	$5 \times 10^{-4}$	0,5	FOCAL	0,4047	0,5029	0,7664
<b>7 ✓</b>	<b>DeiT-B/16</b>	<b>AdamW</b>	<b><math>5 \times 10^{-4}</math></b>	<b>0,5</b>	<b>FOCAL</b>	<b>0,8915</b>	<b>0,8921</b>	<b>0,9700</b>
8	DeiT - B/16	AdamW	$5 \times 10^{-3}$	0,5	FOCAL	0,7752	0,7979	0,9229
9	DeiT - B/16	AdamW	$5 \times 10^{-4}$	0,5	BCE.	0,9058	0,8852	0,9681
10	DeiT - B/16	Adam	$5 \times 10^{-3}$	0,5	BCE.	0,7302	0,7533	0,8957

**Nota.** ✓ Configuración final seleccionada para proveer pesos al modelo híbrido. Experimentos hechos en el NB3-Modelo\_ViT y guardados en experiments\_vit.csv, donde LR = Learning Rate.

### 3.4.2. Entrenamiento

Con base en los resultados de la Tabla 7, se seleccionó el Experimento 7 (DeiT-B/16) como configuración final para la rama ViT. A diferencia de la selección realizada en la CNN, donde el experimento ganador coincidía con la sensibilidad más alta sin sacrificar significativamente las demás métricas, en este caso se priorizó el balance general del modelo. El Experimento 7 supera

al Experimento 10 en cinco de las seis métricas evaluadas, la única métrica en la que el Experimento 10 resulta superior es la sensibilidad (0,9058 vs. 0,8915), con una diferencia de 0,0079, equivalente a menos de un paciente adicional detectado por cada 100 casos malignos.

El modelo final se entrenó con AdamW, Focal Loss como función de pérdida, CosineAnnealingLR como Scheduler y precisión mixta AMP durante 20 epochs con detención anticipada (patience de 5 epochs). El modelo completó las 20 epochs sin que el EarlyStopping se activara, con un tiempo de entrenamiento de aproximadamente 170 minutos. La configuración descrita se expande en la Tabla 8.

**Tabla 8. Configuración final Modelo ViT**

Hiperparámetro	Valor
Backbone	DeiT-B/16
Capa oculta del encabezado	Dropout(0,5) → Linear(768→1)
Número de salidas	1 sigmoide → 2 clases – benign / malignant –
Optimizador	AdamW
Tasa de aprendizaje	$5 \times 10^{-5}$
Weight Decay	$1 \times 10^{-4}$
Epochs	20
Función de pérdida	FOCAL Loss
Planificador LR	CosineAnnealingLR
Acelerador	GPU T4 x2

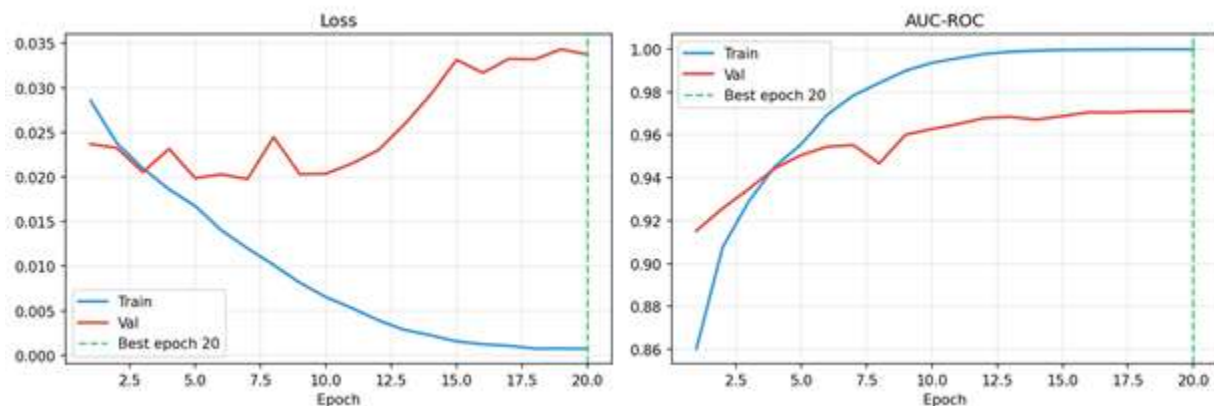
**Nota.** Elaboración a partir de la configuración del notebook 3-modelo-vit.ipynb

### 3.4.3. Resultados

La Figura 8 muestra las curvas de pérdida y AUC-ROC a lo largo del entrenamiento. La dinámica de la pérdida refleja el comportamiento esperado bajo Focal Loss: los valores absolutos son notablemente menores que en BCE (pérdida de entrenamiento final de 0,0007; pérdida de validación final de 0,0361).

#### Figura 8.

Curvas de entrenamiento — ViT DeiT-B/16



**Nota.** Izquierda: pérdida (BCE) de entrenamiento y validación por epoch. Derecha: AUC-ROC de entrenamiento y validación.. La línea punteada verde marca la epoch de mejor AUC de validación – epoch 20 –.

La pérdida de validación muestra una ligera divergencia desde la epoch 7 aproximadamente mientras la de entrenamiento continúa descendiendo de forma continua, sin embargo, pese a esta divergencia en pérdida, el AUC de validación continuó mejorando durante todo el entrenamiento, por lo que esta tendencia no fue un indicador directo de la calidad de clasificación a nivel de umbral variable.

La progresión del AUC de validación fue sostenida y sin retrocesos significativos: 0,9152 (epoch 1) → 0,9627 (epoch 10) → 0,9711 (epoch 20). El AUC de entrenamiento alcanzó

valores próximos a 1,0 desde la epoch 15 en adelante, creando una brecha visible frente al AUC de validación indicando una tendencia de sobreajuste; sin embargo, dado que el AUC de validación continuó aumentando durante todo el entrenamiento, el checkpoint de la epoch 20 representa el mejor estado de generalización disponible.

La Tabla 9 presenta el reporte de clasificación del modelo DeiT-B/16 (Exp. 7) sobre el conjunto de prueba ( $n = 3.800$ ) con umbral  $\tau = 0,5$ .

**Tabla 9. Reporte de clasificación modelo ViT**

Clase	Precisión	Sensibilidad	F1-score	Soporte (Imágenes)
Benigno	0.9367	0.9375	0,9371	2399
Maligno	0.8928	0.8915	0,8921	1401
Exactitud global	—	0,9205	—	3800
AUC - ROC	—	0,9700	—	—

**Nota.** Métricas calculadas con sklearn.metrics.

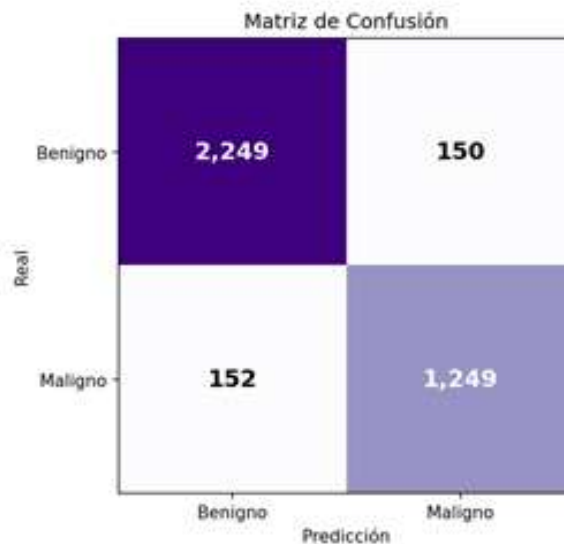
Los resultados del DeiT-B/16 superan a la CNN (EfficientNet-B0) en AUC (0,9700 vs. 0,9628), exactitud (0,9211 vs. 0,9084) y tasa de falsos negativos (10,2% vs. 11,9%). Este resultado es atribuible principalmente a dos factores; el primero, la estrategia de preentrenamiento por destilación de DeiT, que le permite aprender representaciones más ricas en el régimen de datos moderado disponible; y segundo, el uso de Focal Loss el cual para esta arquitectura si mejora el

rendimiento, contrario que en el CNN; esta función de pérdida orienta el aprendizaje hacia los casos de malignidad más difíciles de clasificar.

Entendiendo esto, se evidencian los resultados absolutos de las predicciones dadas por el modelo, interpretadas en la matriz de confusión ilustrada en la Figura 9.

**Figura 9.**

*Matriz de confusión - Modelo ViT*



**Nota.** Matriz de confusión calculada con sklearn.metrics; valores absolutos sobre el conjunto de prueba.

Como se aprecia, de los 3.800 casos del conjunto de prueba, el modelo clasificó correctamente 2.249 de los 2.399 casos benignos y 1.249 de los 1.401 casos malignos acumulando 150 falsos positivos y 152 falsos negativos. Comparado con la CNN —187 falsos positivos y 167 falsos negativos sobre el mismo conjunto—, el ViT reduce ambos tipos de error de forma simultánea: detecta 15 malignos adicionales y genera 37 alarmas falsas menos, equivalentes a aproximadamente 15 biopsias innecesarias evitadas por cada 1.000 pacientes benignos; por lo que este modelo prevé una mejora en las dos direcciones clínicamente relevantes; que el DeiT supere

a la CNN en sensibilidad y especificidad simultáneamente indica una separación entre clases genuinamente superior, atribuible a la representación más rica que aporta el preentrenamiento por destilación combinado con Focal Loss.

### **3.5. Modelo Híbrido CNN + ViT**

Esta sección describe la arquitectura, el entrenamiento y los resultados de la primera iteración del modelo híbrido CNN + ViT, construido sobre los checkpoints de los modelos individuales entrenados en las secciones 3.3 y 3.4.

#### ***3.5.1. Motivación para la fusión de modelos***

El análisis de los errores cometidos por los modelos individuales revela patrones que justifican su fusión. La CNN (AUC = 0,9628) produjo 167 falsos negativos y 187 falsos positivos sobre el conjunto de prueba; el ViT (AUC = 0,9700) produjo 152 falsos negativos y 150 falsos positivos. El hecho de que ambos modelos tengan desaciertos en magnitudes distintas y en casos distintos, es indicativo de que extraen representaciones diferentes de la misma imagen. Esta complementariedad es estructuralmente coherente con las diferencias en el mecanismo de procesamiento: la CNN captura eficientemente patrones texturales y morfológicos como bordes o gradientes de pigmentación. Mientras que el Transformer modela relaciones de largo alcance: coherencia cromática, simetría y contexto perilesional [41].

La fusión se realiza a nivel de representación intermedia (feature fusion o late fusion), concatenando el vector de 1.280 dimensiones producido por el bloque de pooling de la CNN con el vector de 768 dimensiones del token del ViT, generando un vector fusionado de 2.048 dimensiones. Esta estrategia permite que el módulo de integración aprenda la ponderación óptima

de cada fuente en función del contenido de la imagen específica, en lugar de imponer una combinación fija o predefinida [42].

### 3.5.2. *Arquitectura del módulo de fusión*

El modelo híbrido está compuesto por tres bloques funcionales secuenciales: los backbones especializados, la compuerta de atención y el clasificador final. Los backbones se inicializan a partir de los mejores checkpoints de NB02 (CNN) y NB03 (ViT), lo que les confiere especialización previa en imágenes dermatoscópicas antes del ajuste conjunto. Solo se conservan los módulos de extracción de características de cada backbone —eliminando los encabezados de clasificación individuales—, de modo que cada rama contribuye con su representación intermedia al módulo de fusión.

La compuerta de atención proyecta linealmente el  $f_c = [f_{CNN}; f_{ViT}] \in \mathbb{R}^{2048}$  y aplica activación sigmoide elemento a elemento siguiendo la ecuación:

$$g = \sigma (W_g * f_c + b_g),$$

donde  $f_c$  es el vector fusionado,  $W_g$  es la matriz de pesos aprendidos por los modelos individuales —con  $W_g \in \mathbb{R}^{2048 \times 2048}$ — y  $b_g$  representa el sesgo (bias).

El resultado es un vector de pesos  $g \in (0, 1)^{2048}$  que actúa como máscara continua por dimensión, produciendo el vector modulado  $f_h = g \odot f_c$ . Esta operación permite al modelo suprimir dimensiones de baja relevancia diagnóstica y amplificar aquellas más discriminativas, adaptando dinámicamente la ponderación a cada imagen.

El clasificador final procesa  $f_h$  mediante LayerNorm  $\rightarrow$  Dropout  $\rightarrow$  Linear(2048, 512)  $\rightarrow$  GELU  $\rightarrow$  Dropout(0,4)  $\rightarrow$  Linear(512, 1), produciendo un único logit cuya sigmoide representa la probabilidad de malignidad. La Tabla 10 resume los componentes del modelo.

**Tabla 10. Componentes y parámetros del modelo híbrido**

Componente	Descripción	Dimensión de salida	Parámetros
CNN Backbone	Características locales; pesos NB2	1.280	4.008.829
ViT Backbone	Características globales; token [CLS]; pesos NB3	768	85.799.425
Concatenación	$f_c = [f_{CNN}; f_{ViT}]$	2.048	0
Attention Gate	Linear(2048, 2048) → Sigmoid	2.048	Aprox 4.198.656
Clasificador final	LayerNorm → DO(0,4) → Linear(2048,512) → GELU → DO(0,4) → Linear(512,1)	1 (logit)	Aprox 1.051.393
<b>Total del modelo</b>	—	—	<b>95.058.303</b>

*Nota.* Parámetros verificados mediante `model.parameters()` en NB04. El backbone ViT corresponde al checkpoint ViT-B/16 de NB03. DO = Dropout.

### 3.5.3. Entrenamiento

Dado que los backbones provienen de modelos ya entrenados en imágenes dermatoscópica, el espacio de hiperparámetros relevante para el híbrido es sustancialmente más reducido que el de los modelos individuales debido a que los parámetros de los backbones son heredados. La única dimensión de ajuste no trivial corresponde al nivel de regularización del clasificador de fusión, explorada mediante el dropout.

Para esto, se realizaron dos experimentos probando dos configuraciones distintas de dropout, manteniendo fijos todos los demás hiperparámetros, que son: ( $lr\_backbone = 5 \times 10^{-6}$ ,  $lr\_head = 1 \times 10^{-4}$ ),  $weight\_decay = 1 \times 10^{-4}$ , Focal Loss como función de pérdida,

CosineAnnealingLR , precisión mixta AMP y detención anticipada con  $\text{patience} = 5$ . Los resultados se evidencian en la Tabla 11.

**Tabla 11. Experimentación Modelo Híbrido**

Ex p	Dropout	AUC - ROC	Sensibilidad	Especificidad	Precisión	F1	Tiempo total
1	0,5	0,9731	0,8794	0,9521	0,9146	0,8967	159,4 min
2	0,3	0,9731	0,9101	0,9275	0,8799	0,8947	87,3 min

**Nota.** Resultados sobre el conjunto de prueba. Tiempo total de entrenamiento en GPU Tesla T4 x2.

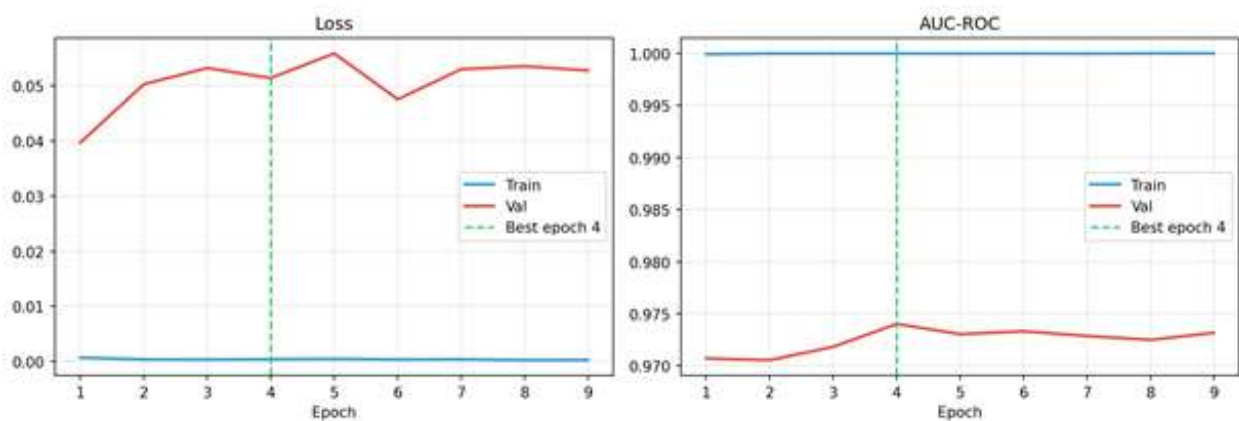
Ambas configuraciones alcanzaron un AUC idéntico de 0,9731, lo que indica que el rendimiento discriminativo del modelo híbrido es robusto al nivel de regularización en el rango explorado. Sin embargo la diferencia principal reside en el balance sensibilidad-especificidad: el Experimento 1 ( $\text{dropout} = 0,50$ ) exhibe mayor especificidad (0,9521 vs. 0,9275) y precisión (0,9146 vs. 0,8799), mientras que el Experimento 2 ( $p = 0,30$ ) logra una sensibilidad superior en 3,1 puntos porcentuales (0,9101 vs. 0,8794). Como se explicó anteriormente, la métrica de sensibilidad —recall de la clase maligna— es la métrica de mayor relevancia clínica para esta aplicación, pues un falso negativo implica dejar sin detectar una lesión potencialmente maligna. Con base en este criterio, se seleccionó el Experimento 2 ( $p = 0,30$ ) como configuración final del modelo híbrido.

### 3.5.4. Resultados

**3.5.4.1 Resultados del entrenamiento.** La Figura 10 muestra las curvas de entrenamiento del modelo seleccionado. El panel de pérdida (Loss) revela un colapso inmediato de la pérdida de entrenamiento hacia cero desde la primera epoch, mientras la pérdida de validación parte en 0,040, alcanza un mínimo en la epoch 4 ( $\sim 0,044$ ) y sube gradualmente hasta  $\sim 0,052$  en la epoch 9; brecha característica de un modelo con alta capacidad relativa (95M parámetros una vez integrados los pesos CNN y ViT) entrenando sobre un dataset finito. Sin embargo, como se detalla más adelante, la detención anticipada en la epoch 9 evita que esta divergencia se traduzca en pérdida de generalización. El panel AUC-ROC confirma esto: el AUC de entrenamiento se mantiene en 1,000 de forma plana desde la epoch 1, mientras el AUC de validación exhibe la progresión  $0,9708 \rightarrow 0,9718 \rightarrow 0,9740$  (óptimo, epoch 4)  $\rightarrow 0,9730 \rightarrow 0,9733 \rightarrow 0,9731$ , observándose una oscilación en el rango  $[0,970 - 0,974]$ .

#### Figura 10.

*Curvas de entrenamiento — Modelo Híbrido CNN + ViT (Exp. 2, dropout = 0,30)*



**Nota.** Izquierda: Focal Loss; la línea verde discontinua marca la epoch 4 (óptima). Derecha: AUC-ROC de entrenamiento y validación (pico 0,9740 en epoch 4).

**3.5.4.2 Resultados de métricas e intervalo de confianza.** La aplicación del umbral óptimo de Youden ( $\tau^* = 0,4358$ ) mejora adicionalmente la sensibilidad hasta 0,9208, identificando un 92,1 % casos malignos a costa de una leve reducción de la especificidad a 0,9221. La posición del umbral óptimo (0,4358) es coherente con la distribución de probabilidades del modelo: las predicciones se concentran en los extremos del espectro de probabilidad, lo que permite al índice de Youden localizar un punto de operación donde la ganancia marginal en sensibilidad compensa exactamente la pérdida marginal en especificidad. En la Tabla 12 se muestran dichos valores justo al intervalo de confianza obtenido por métrica evaluada.

**Tabla 12. Resultados de valores obtenidos vs IC**

Métrica	Valor (Umbral 0.5)	Intervalo de confianza (95%)
AUC-ROC	0,9731	[0,9681 – 0,9777]
Sensibilidad (recall maligno)	0,9101	[0,8938 – 0,9248]
Especificidad (recall benigno)	0,9275	[0,9168 – 0,9378]
Precisión	0,8799	[0,8614 – 0,8960]
F1	0,8947	[0,8816 – 0,9067]
Exactitud	0,9211	—

**Nota.** Los IC se obtuvieron mediante remuestreo bootstrap ( $B = 1.000$ ) sobre el conjunto de prueba.

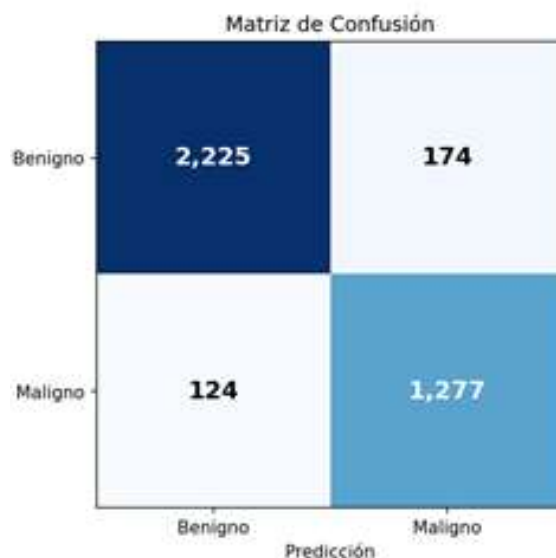
El AUC-ROC de 0,9731 [IC 95%: 0,9681 – 0,9777] sitúa al modelo híbrido por encima de ambos modelos individuales: CNN (0,9628) y DeiT-B/16 (0,9700). La mejora de 1,03 puntos porcentuales sobre el DeiT —el mejor modelo individual— es estadísticamente robusta. Este resultado verifica la hipótesis central de la arquitectura propuesta: la fusión de representaciones

locales (CNN) y globales (ViT) mediante una compuerta de atención aprendida permite superar el rendimiento de cualquiera de las ramas individuales.

**3.5.4.3. Resultados absolutos.** La matriz de confusión ilustrada en la Figura 11 muestra los resultados absolutos sobre las 3.800 imágenes del conjunto de validación, en estos se obtienen 1.277 malignos correctamente identificados (VP, 91,1 %), 124 malignos no detectados (FN, 8,9 %), 2.225 benignos correctamente clasificados (VN, 92,7 %) y 174 benignos predichos erróneamente como malignos (FP, 7,3 %).

**Figura 11.**

*Evaluación clínica — Modelo Híbrido CNN + ViT*



**Nota.** Derecha: matriz de confusión donde VN = 2.225, FP = 174, FN = 124, VP = 1.277.

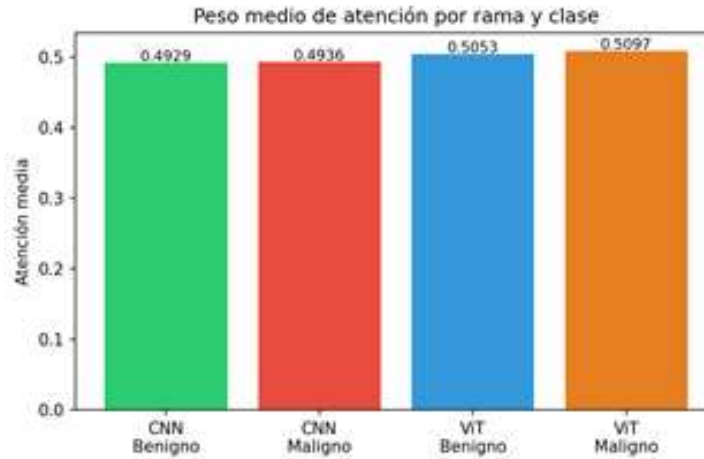
En términos comparativos, los 124 FN del híbrido representan una reducción de 43 casos respecto a la CNN (167 FN) y 28 casos respecto al DeiT (152 FN), traduciendo directamente la ganancia en sensibilidad en menos lesiones malignas no detectadas sobre el mismo conjunto de prueba.

**3.5.4.4. Attention gate.** Para analizar el comportamiento del módulo de fusión, se estudiaron los pesos de atención definidos como  $g = \sigma (W_g * f_h )$ , sobre el conjunto de prueba completo. Dado que  $g \in \mathbb{R}^{2048}$ , sus componentes se dividieron en dos subconjuntos correspondientes a cada rama del modelo: los primeros 1.280 elementos asociados a la representación extraída por la CNN y los 768 restantes a la representación del ViT.

La ilustración de la Figura 12 muestra que la rama CNN recibe un peso medio de 0,4929 para casos benignos y 0,4936 para malignos ( $\Delta_{\text{CNN}} = 0,0007$ ), mientras que la rama ViT recibe 0,5053 para benignos y 0,5097 para malignos ( $\Delta_{\text{ViT}} = 0,0044$ ). Primero, la rama ViT recibe pesos medios ligeramente superiores a la CNN en ambas clases ( $\sim 0,51$  vs.  $\sim 0,49$ ), lo cual es coherente con que el DeiT-B/16 individual alcanzó  $\text{AUC} = 0,9700$  frente a 0,9628 de la CNN: la compuerta codifica de forma implícita la mayor capacidad discriminativa del backbone Transformer. Segundo, el diferencial inter-clase es seis veces mayor en la rama ViT ( $\Delta = 0,0044$ ) que en la CNN ( $\Delta = 0,0007$ ), indicando que la compuerta asigna una ponderación marginalmente más alta a la representación ViT precisamente en los casos malignos, donde las características globales de largo alcance son más discriminativas.

**Figura 12.**

*Peso medio de atención por rama según clase diagnóstica*

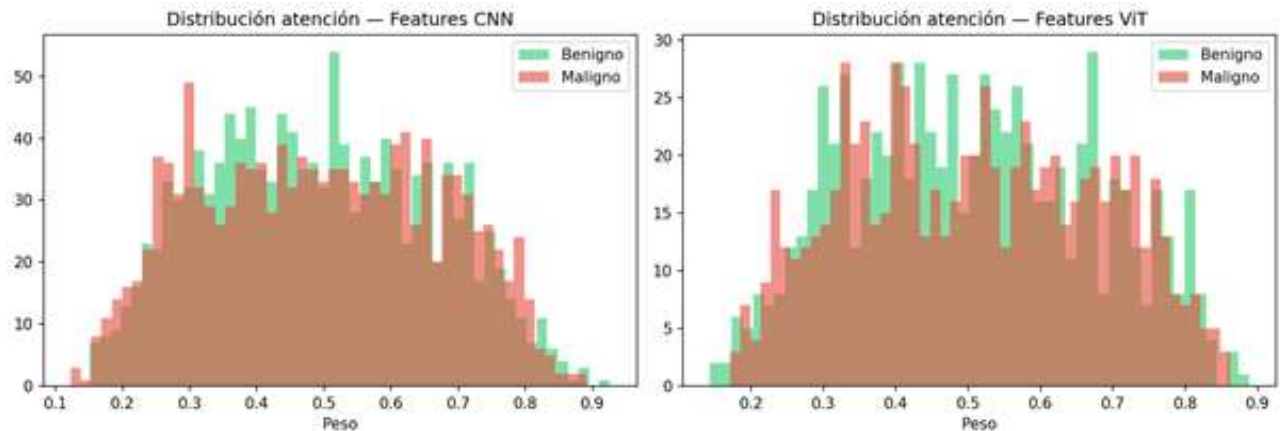


*Nota.* Los valores se obtienen promediando las dimensiones del vector  $g$  correspondientes a cada backbone.

En cuanto a la distribución de los pesos individuales por rama, la Figura 13 ilustra los paneles de distribución de estos, donde se revela un patrón estructuralmente distinto entre ambas ramas. La distribución de pesos CNN (panel central) forma una campana amplia centrada en  $\sim 0,5$ , con pesos que se distribuyen entre 0,10 y 0,90 y una superposición prácticamente total entre las distribuciones de benignos y malignos; la separación entre clases es mínima y estadísticamente consistente con los valores medios observados.

**Figura 13.**

*Peso individual de atención por rama*



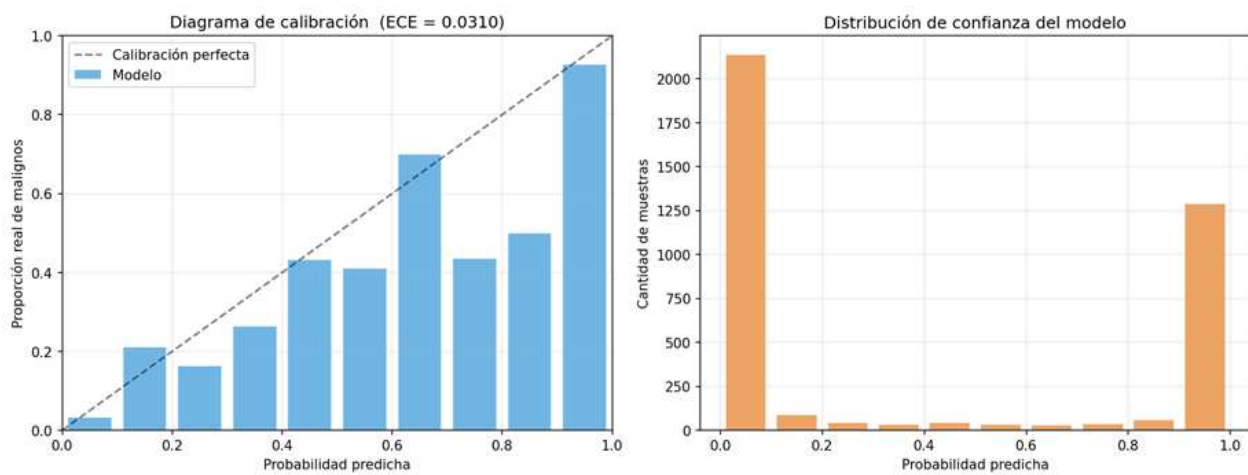
*Nota.* En la izquierda: distribución de pesos CNN por clase (benigno en verde, maligno en rojo). En la Derecha: distribución de pesos ViT por clase, con leve mayor separación inter-clase que CNN.

La distribución ViT (panel derecho) es más compacta, concentrada en el rango [0,20 – 0,85], con leve asimetría positiva hacia valores altos; la separación entre clases es igualmente pequeña pero ligeramente más pronunciada que en CNN, consistente con el mayor  $\Delta$  observado. En ambas ramas, la proximidad de los pesos medios a 0,5 indica que la compuerta opera cerca de la región lineal de la sigmoide, sin producir supresiones fuertes en ninguna dimensión.

**3.5.4.5. Calibración Probabilística.** La calibración del modelo híbrido se evaluó mediante el Expected Calibration Error (ECE) y el diagrama de fiabilidad (reliability diagram), siguiendo la metodología de Guo et al. [36]. En la siguiente Figura 14 se presentan los dos componentes del análisis: el diagrama de calibración y la distribución de confianza del modelo.

**Figura 14.**

*Calibración probabilística — Modelo Híbrido CNN + ViT*



**Nota.** En la Izquierda: diagrama de fiabilidad ; cada barra indica la proporción real de casos malignos en el intervalo de probabilidad predicha correspondiente; la línea discontinua es la calibración perfecta. En la derecha: distribución de confianza del modelo.

El diagrama de calibración muestra la proporción real de casos malignos por intervalos de probabilidad predicha frente a la diagonal de calibración perfecta. El modelo está bien calibrado en los extremos. En los intervalos intermedios existen desviaciones, pero con alta varianza debido al bajo número de muestras.

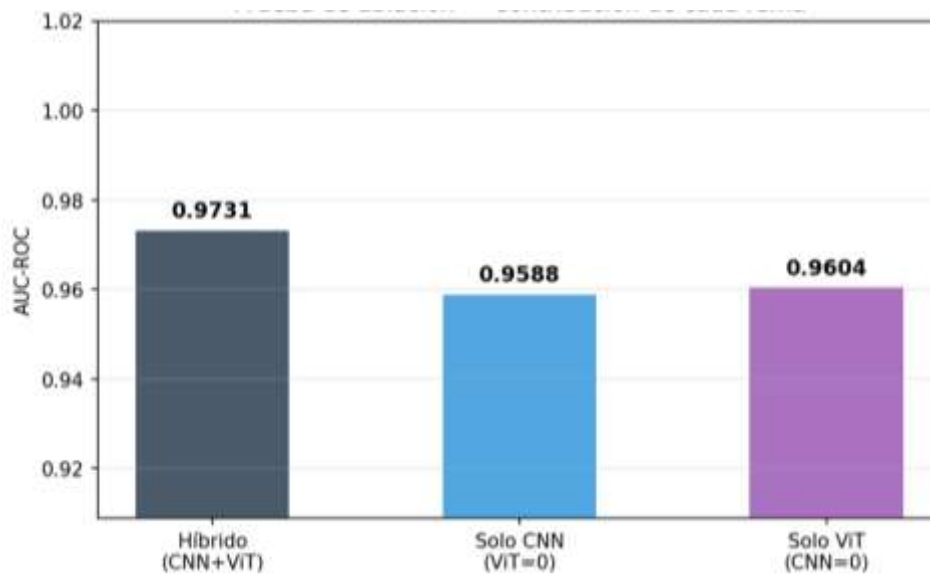
El ECE global es 0.031, indicando excelente calibración ( $< 0.05$ ) según el umbral estándar de la literatura. El bajo ECE se explica por la distribución de predicciones: la mayoría de las muestras se concentra en los extremos ( $[0.0-0.1]$  y  $[0.9-1.0]$ ), mientras que los intervalos

intermedios contienen pocos casos. Dado que el ECE pondera por frecuencia, los bins centrales tienen impacto mínimo, y los extremos dominan el resultado.

**3.5.4.6. Estudio de ablación.** Para cuantificar la contribución independiente de cada rama al rendimiento del modelo híbrido, se realizó un estudio de ablación anulando el vector de características de un backbone —sustituyéndolo por un vector de ceros— mientras se mantenían los pesos del modelo completo entrenado. Esta metodología permite aislar el efecto de cada rama sobre la salida del attention gate y del clasificador sin reentrenar el modelo. Los resultados se ilustran en la Figura 15.

**Figura 15.**

*Estudio de ablación – contribución de cada rama al AUC-ROC del modelo híbrido.*



*Nota.* Comparación del AUC- ROC de los tres modelos. La ablación se realizó sobre el conjunto de prueba completo ( $n = 3.800$ ) sin reentrenamiento.

La eliminación de la rama ViT (Solo CNN, ViT = 0) produce una caída de 1,43 pp (AUC = 0,9588), mientras que la eliminación de la CNN (Solo ViT, CNN = 0) produce una caída de 1,27 pp (AUC = 0,9604).

La caída ligeramente mayor al eliminar el ViT (-1,43 pp vs. -1,27 pp de la CNN) es coherente con el análisis del attention gate: la rama ViT exhibe un mayor diferencial de pesos entre clases ( $\Delta_{ViT} = 0,0044$  vs.  $\Delta_{CNN} = 0,0007$ ), lo que indica que el DeiT contribuye con señales cualitativamente más discriminativas entre benigno y maligno.

### **3.6. Desarrollo del prototipo web**

Para llevar a cabo el desarrollo del prototipo web se realizó un análisis detallado de los requisitos necesarios para garantizar tanto la utilidad clínica como la viabilidad técnica de la plataforma de apoyo al diagnóstico dermatológico. Este proceso se fundamentó en tres ejes principales:

1. Análisis del dominio clínico e investigativo: se evaluaron los flujos de trabajo tradicionales en el diagnóstico de lesiones cutáneas y se consultó literatura sobre la integración de modelos de aprendizaje profundo en la práctica médica. Esto permitió identificar los pasos críticos que debía contemplar el prototipo, tales como la gestión del historial fotográfico del paciente y la validación obligatoria del diagnóstico sugerido por la Inteligencia Artificial por parte del médico especialista.
2. Evaluación de la arquitectura tecnológica: Se realizó un análisis integral de las capacidades y limitaciones del entorno web basado en una arquitectura cliente-servidor. Este proceso incluyó la evaluación del ecosistema de React para el desarrollo y despliegue de la interfaz de usuario, la eficiencia y escalabilidad de FastAPI para la exposición del modelo

predictivo híbrido (CNN–ViT) como servicio, y los mecanismos de seguridad y gestión de datos ofrecidos por Supabase como plataforma Backend as a Service (BaaS).

Este análisis permitió identificar restricciones clave relacionadas con el procesamiento de imágenes, los tiempos de respuesta del sistema (latencia) y la protección de datos sensibles, las cuales fueron consideradas como criterios fundamentales en el diseño e implementación del prototipo.

3. Definición funcional: A partir de la integración de los elementos analizados previamente, se definió un conjunto de requerimientos funcionales y no funcionales que guiaron el diseño de la base de datos, la interfaz de usuario y los mecanismos de comunicación de la API (interfaz de programación de aplicaciones).

En las siguientes tablas se presentan los requisitos técnicos y operativos del sistema, organizados en las categorías de requerimientos funcionales y no funcionales.

### ***3.6.1 Requerimientos funcionales***

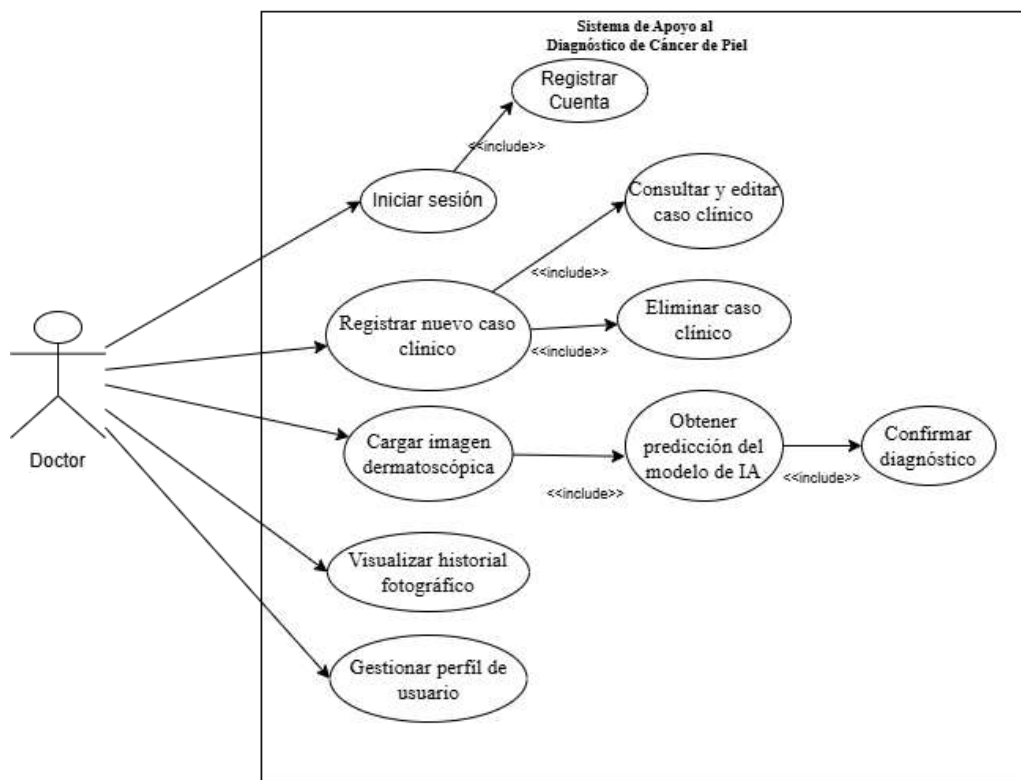
Los requerimientos funcionales describen el comportamiento observable del sistema en respuesta a las acciones del usuario. Cada requerimiento se identificó a partir del análisis de los procesos clínicos involucrados en el diagnóstico dermatológico asistido por inteligencia artificial, priorizando aquellas funcionalidades directamente relacionadas con la gestión de pacientes, el análisis de imágenes y la visualización de resultados diagnósticos. (Véase las tablas de requerimientos funcionales y no funcionales en el Apéndice A)

### 3.6.2 Casos de uso

Los casos de uso describen las interacciones entre los actores del sistema y las funcionalidades que este provee. El actor principal identificado es el especialista médico (dermatólogo o médico general), quien opera la plataforma para el diagnóstico asistido de lesiones cutáneas. Los casos de uso se ilustran en la Figura 16, siendo estos derivados directamente de los requerimientos funcionales establecidos en la sección 3.6.1. (Véase las tablas en el Apéndice A)

**Figura 16.**

*Diagrama de Casos de Uso*



*Nota.* Diagrama de descripción de los casos de uso de los usuarios del sistema

### 3.6.3 Diseño del sistema

El diseño del sistema se concibió con el objetivo de integrar de manera coherente los distintos componentes que conforman el prototipo de apoyo al diagnóstico de cáncer de piel. Este

diseño responde a los requerimientos funcionales y no funcionales previamente definidos, garantizando una interacción eficiente entre la interfaz web, el modelo predictivo basado en aprendizaje profundo y los servicios de almacenamiento y procesamiento. Asimismo, busca asegurar un flujo adecuado de la información, desde la captura de imágenes hasta la generación de resultados, manteniendo criterios de rendimiento, escalabilidad y seguridad en el manejo de datos.

### ***3.6.4 Diseño de la base de datos***

Para la gestión de la persistencia, se implementó un sistema de base de datos relacional utilizando PostgreSQL a través de la infraestructura de Supabase. La elección de este modelo responde a la necesidad de mantener relaciones estrictas y consistentes entre la información demográfica del paciente y sus múltiples registros de diagnóstico, garantizando la integridad de los datos y la trazabilidad histórica de cada caso. En este contexto, el modelo entidad-relación se fundamenta en una relación de uno a muchos (1:N) entre las entidades de pacientes y diagnósticos, permitiendo que un único paciente esté asociado a una secuencia de análisis a lo largo del tiempo, lo que facilita el seguimiento evolutivo de las lesiones dermatológicas. Las Tablas 13 y 14 corresponden a la estructura descrita de tablas en la base de datos.

***Tabla 13. Pacientes***

<b><i>Campo</i></b>	<b><i>Tipo de dato</i></b>	<b><i>Descripción</i></b>
id	UUID	Identificador único del paciente
nombre	Varchar	Nombres del paciente
apellido	Varchar	Apellidos del paciente
color_piel	Varchar	Clasificación según escala

		clínica.
estado	Varchar	Estado actual del registro (Diagnosticado/Sin diagnóstico).
user_id	UUID	Referencia al médico propietario del registro

*Nota.* Esta tabla almacena la información base del sujeto de estudio y permite la segmentación de datos por especialista.

**Tabla 14. Diagnóstico**

<i>Campo</i>	<i>Tipo de dato</i>	<i>Descripción</i>
id	UUID	Identificador único del diagnóstico.
paciente_id	UUID	Relación con la tabla Pacientes.
imagen_url	Varchar	Enlace al objeto almacenado en el Storage.
clasificacion	Varchar	Resultado de la IA (Maligno/Benigno).
probabilidad	Float	Nivel de confianza del modelo
confirmado	Boolean	Validación final por parte del especialista

*Nota.* Esta tabla custodia los resultados generados por el modelo híbrido y las validaciones del médico.

Posteriormente, se ilustra en la Figura 17 el esquema visual de la estructura construida para la base de datos relacional en Supabase.

**Figura 17.***Esquema visual de la base de datos*

*Nota.* Supabase gestiona la autenticación nativamente en `auth.users`, vinculándose a los pacientes mediante una llave foránea.

### 3.6.5 Diseño del backend

El backend del sistema se diseñó como una API RESTful implementada en Python bajo el framework FastAPI. A diferencia de las arquitecturas monolíticas tradicionales, este componente se estructuró como un microservicio independiente cuya función exclusiva es actuar como el motor de inferencia analítica de la plataforma. Su objetivo principal es recibir las referencias de las imágenes clínicas, procesarlas mediante el modelo de Deep Learning híbrido (CNN-ViT) y devolver una predicción probabilística en tiempo real a la aplicación cliente, delegando la persistencia de datos a la infraestructura de Supabase.

El desarrollo del backend integró las siguientes tecnologías principales:

- FastAPI: para la definición de rutas, el manejo asíncrono de solicitudes HTTP y la generación automática de la documentación de la API.
- PyTorch: como librería principal de Machine Learning, encargada de cargar los pesos del modelo pre entrenado (.pt.zip), procesar las imágenes como tensores y ejecutar las operaciones de la red neuronal.
- Pydantic: para la validación y serialización de datos, garantizando que los objetos de entrada y salida cumplan con esquemas tipados.
- Requests / HTTPX: para la descarga en memoria de las imágenes almacenadas en el Storage de Supabase.
- Uvicorn: como servidor ASGI para la ejecución en producción de la aplicación FastAPI, permitiendo la concurrencia en las peticiones.

Para garantizar la fiabilidad del sistema en un entorno de salud, la arquitectura del backend incorpora un mecanismo de fail-fast durante su inicialización. Al iniciar el servidor, se verifica la integridad del archivo de pesos del modelo y se extrae su umbral óptimo de decisión. En caso de que el archivo no exista o esté corrupto, el sistema detiene su ejecución.

La API expone dos endpoints principales:

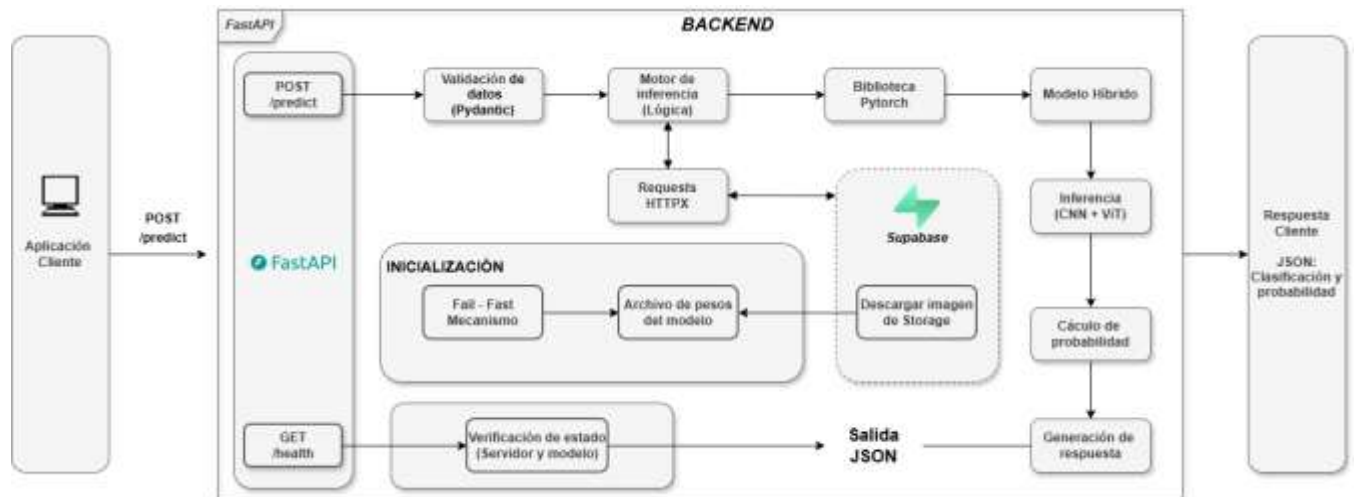
1. **GET /health:**

- *Entrada:* solicitud HTTP sin parámetros.
- *Proceso:* verifica que el servidor esté en ejecución y que el modelo esté cargado en memoria.
- *Salida:* objeto JSON que confirma el estado de la API.

## 2. POST /predict

- Entrada: recibe un documento JSON (PeticionImagen) con la URL pública de la imagen dermatoscópica.
- Proceso:
  1. Descarga la imagen desde la URL proporcionada.
  2. Aplica transformaciones de preprocesamiento (redimensionamiento y normalización).
  3. Ejecuta la inferencia mediante la arquitectura híbrida (EfficientNet + Vision Transformer) sin cálculo de gradientes
  4. Calcula la probabilidad de malignidad mediante una función sigmoide y la compara con el umbral de decisión.
- Salida: devuelve un JSON con la clasificación (“Maligno” o “Benigno”) y la probabilidad.

Como se ilustra en la Figura 18, la arquitectura del backend sigue un enfoque desacoplado orientado a servicios, donde el flujo de inferencia se ejecuta de manera independiente del cliente y de la capa de almacenamiento.

**Figura 18.***Arquitectura del backend*

*Nota. Arquitectura del Backend de la Aplicación y las tecnologías usadas*

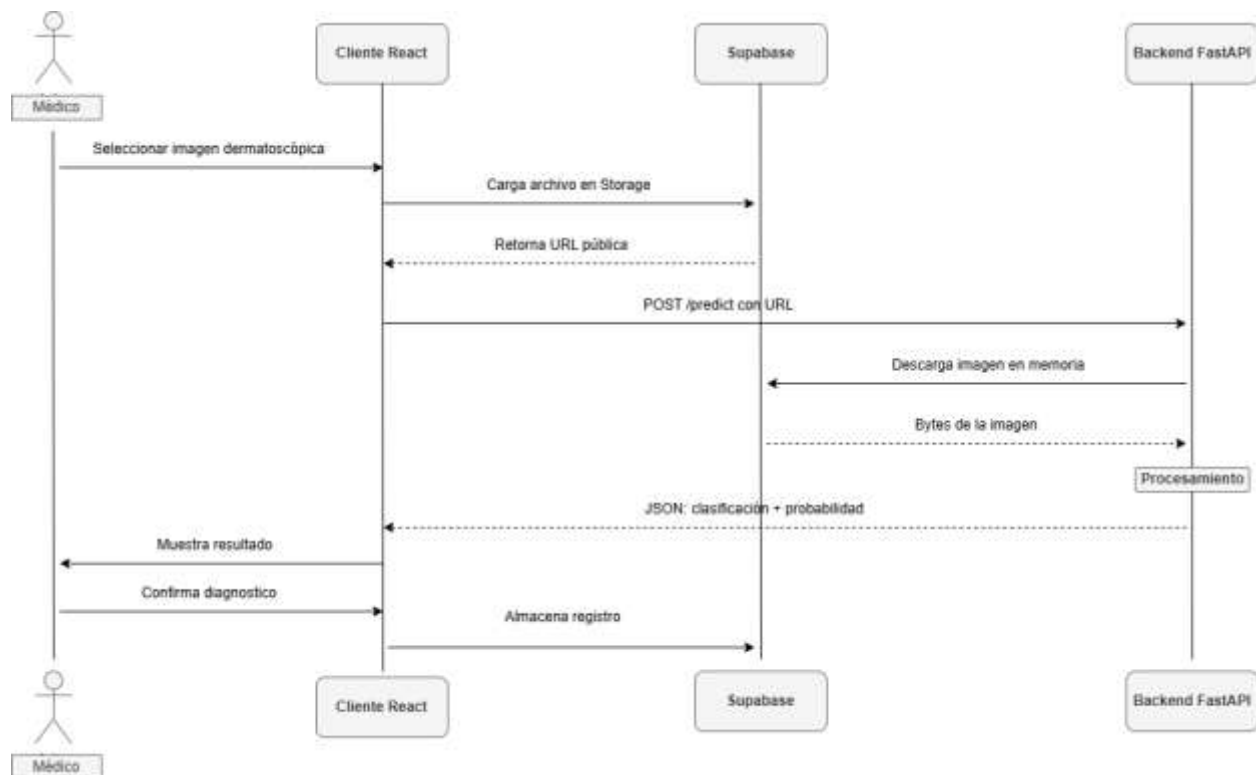
El flujo de comunicación entre la aplicación cliente (React), la base de datos (Supabase) y el backend sigue la secuencia que se presenta en la Figura 19:

1. El especialista selecciona una imagen desde la interfaz.
2. La aplicación cliente carga el archivo en el Storage de Supabase y obtiene una URL pública.
3. El cliente envía la URL al backend mediante una petición POST al endpoint /predict.
4. El backend descarga la imagen en memoria, ejecuta la inferencia y retorna el resultado.
5. La interfaz actualiza la visualización y permite confirmar el diagnóstico.
6. Finalmente, el cliente construye el registro completo (datos del paciente, imagen y resultado) y lo almacena en la base de datos.

De esta manera, el backend garantiza un procesamiento eficiente y desacoplado, evitando la transferencia innecesaria de archivos binarios pesados y concentrando los recursos computacionales en la inferencia del modelo predictivo.

**Figura 19.**

*Diagrama de secuencia*



*Nota.* Diagrama de secuencia entre el usuario de la aplicación y de las partes que la conforman.

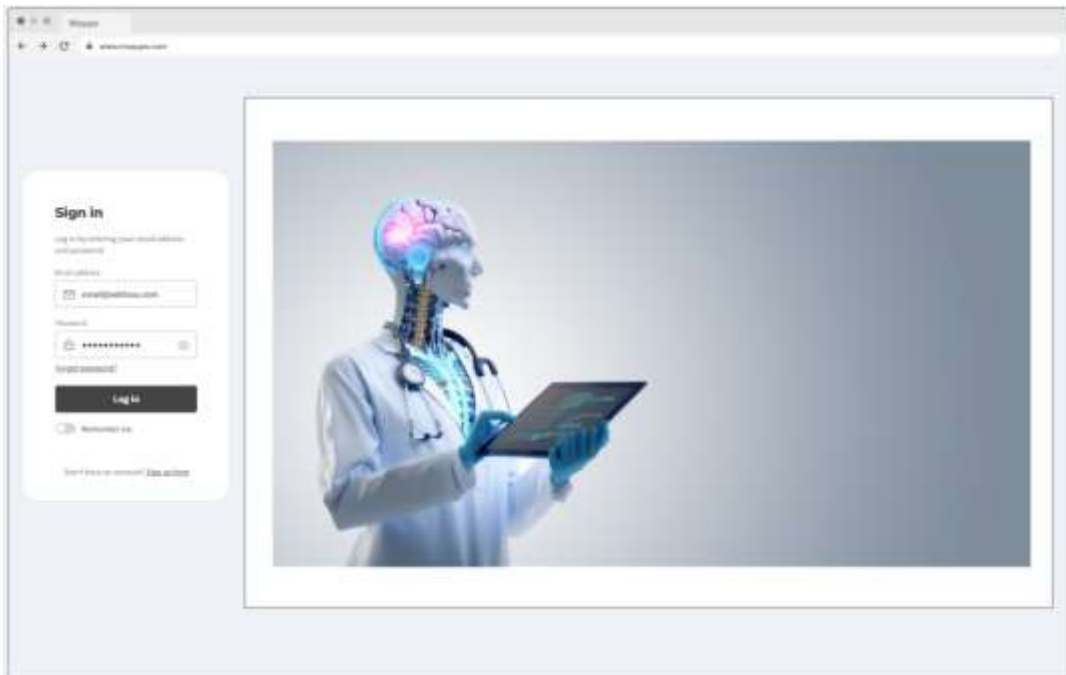
### 3.6.6 Diseño del frontend

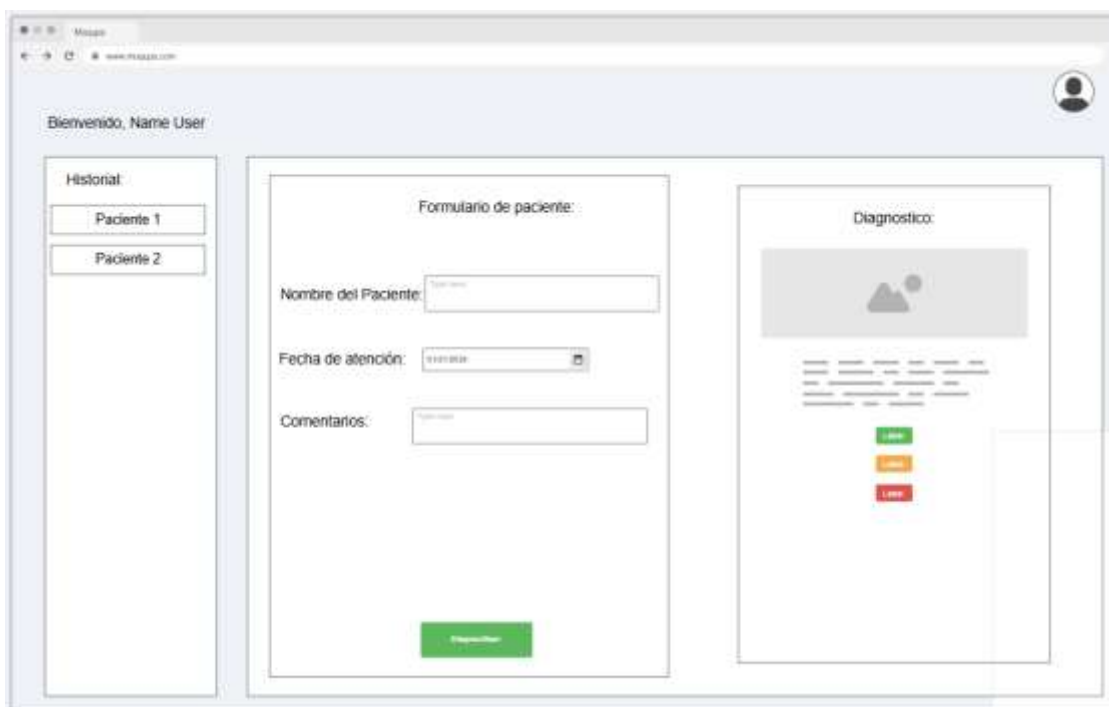
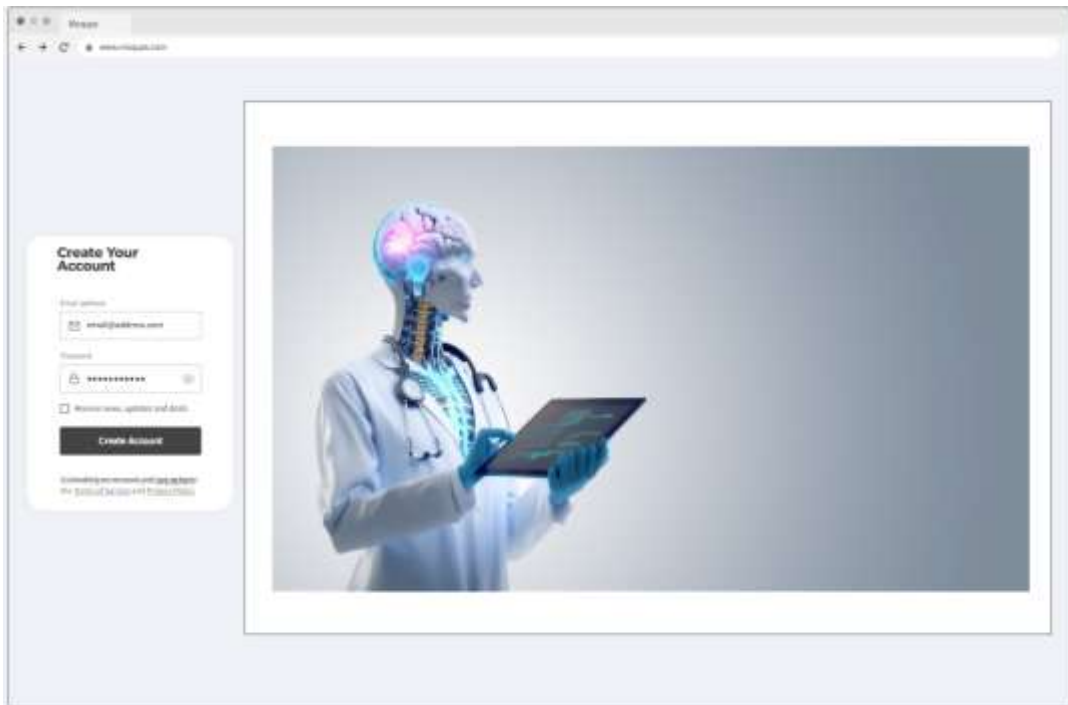
El frontend del sistema constituye la capa de interacción entre el especialista médico y las capacidades de clasificación del modelo híbrido. Su diseño responde directamente a los

requerimientos funcionales y no funcionales establecidos, priorizando la usabilidad clínica (RNF02), la responsividad en dispositivos móviles y la claridad en la presentación de los resultados diagnósticos.

Como se presenta en la Figura 20, se desarrollaron mockups que definen la estructura visual y la organización de los componentes de la interfaz, sirviendo como base para la implementación final del sistema.

**Figura 20.**  
*Mockups del sistema*





*Nota. Mockups del sistema según los requerimientos del usuario*

### ***3.6.7 Tecnologías y arquitectura de la interfaz***

El frontend fue desarrollado con React 18 utilizando Vite como entorno de construcción, lo que permite tiempos de compilación reducidos y recarga en caliente durante el desarrollo. El sistema de estilos se implementó con Tailwind CSS, adoptando un enfoque de diseño basado en clases utilitarias que facilita la adaptación responsiva de los componentes mediante breakpoints predefinidos (md para tableta y lg para escritorio), garantizando la correcta visualización de la interfaz tanto en dispositivos móviles como de escritorio en cumplimiento del requerimiento RNF02.

La gestión del estado global de la aplicación se implementó mediante la Context API de React, sin dependencias externas de manejo de estado como Redux o Zustand. Esta decisión responde a la escala del prototipo y facilita el mantenimiento del código. Se definieron dos contextos principales: AuthContext, que centraliza el estado de autenticación del usuario y las operaciones de inicio y cierre de sesión; y CasosContext, que gestiona la lista de casos clínicos, el caso actualmente seleccionado y las funciones de creación, edición, eliminación y actualización diagnóstica.

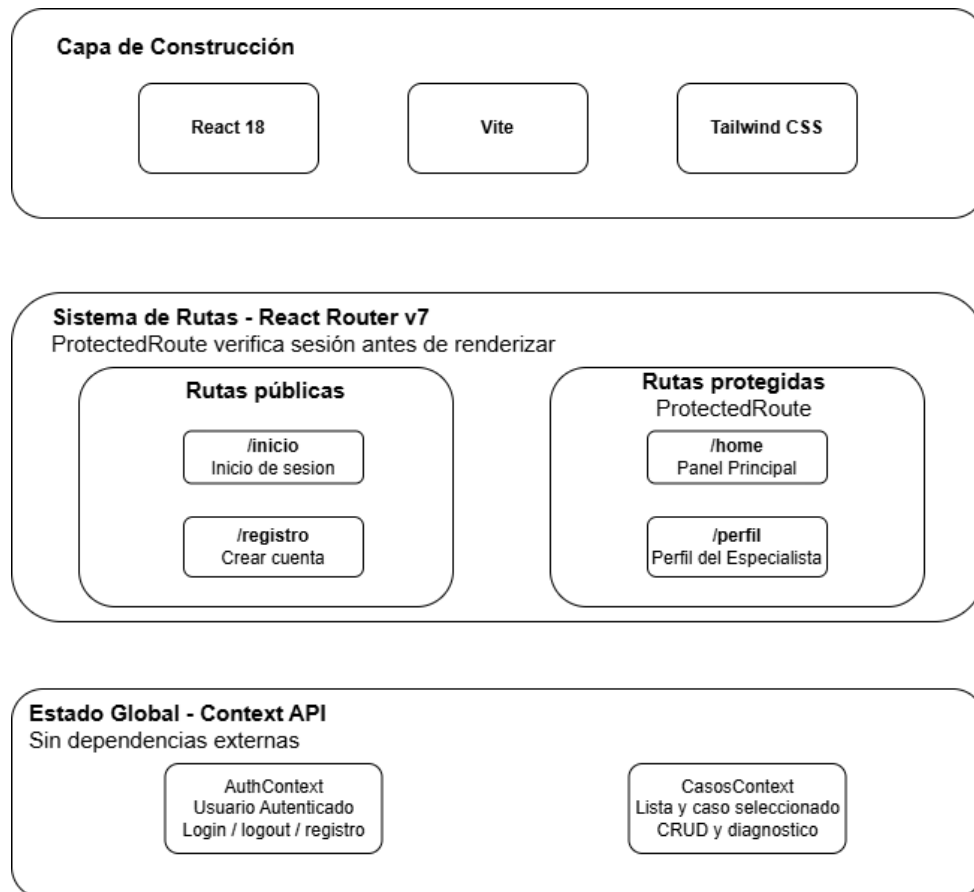
El enrutamiento de la aplicación se gestiona con React Router v7, que diferencia entre rutas públicas (inicio de sesión y registro) y rutas protegidas (panel principal y perfil de usuario). El componente ProtectedRoute verifica la existencia de un usuario autenticado en el contexto antes de renderizar las páginas privadas; en ausencia de sesión activa, presenta una pantalla de acceso restringido con redirección al módulo de inicio de sesión.

Como se presenta en la Figura 21, la arquitectura del frontend y las tecnologías empleadas se organizan en capas que integran el sistema de construcción, la gestión de estado, el enrutamiento

y el diseño responsivo, evidenciando la interacción entre los distintos componentes que conforman la interfaz de usuario.

### Figura 21.

#### *Arquitectura y Tecnologías del Frontend*



*Nota.* Arquitectura de la aplicación en su capa de frontend usando las diferentes tecnologías

### 3.6.8 Planificación de módulos y estructura de la interfaz

La interfaz se organiza en cuatro módulos principales, cada uno implementado como un componente React independiente y comunicado con los demás a través de los contextos globales. En la Tabla 15 se detalla esta organización modular la cual facilita el mantenimiento del código y la incorporación futura de nuevas funcionalidades.

**Tabla 15. Módulos del frontend y requerimientos funcionales asociados**

Módulo	Componente	RF asociados	Descripción
Historial de casos	Sidebar.jsx	RF02	Panel lateral que lista los primeros 15 casos del especialista con scroll interno. Permite seleccionar un caso para visualizar su información.
Ficha del paciente	UploadForm.jsx	RF02, RF07	Formulario central con los datos demográficos y clínicos del paciente. Soporta modos de visualización, edición y creación de nuevos casos.
Panel de diagnóstico	DiagnosisPanel.jsx	RF03, RF04, RF05, RF06, RF07	Panel dedicado al análisis dermatológico: carga de imágenes, historial fotográfico con miniaturas, visualización de clasificación y barra de probabilidad, y confirmación del diagnóstico por el especialista.
Autenticación	InicioPage.jsx RegistroPage.jsx	/ RF01	Páginas de inicio de sesión y registro con validación de formularios, gestión de errores y simulación de latencia de red.
Perfil de usuario	PerfilPage.jsx	RF08	Página de visualización y edición de los datos del especialista, con opción de cierre de sesión.
Navegación	Navbar.jsx	RF01, RF08	Barra de navegación superior con menú desplegable de usuario que expone las acciones de perfil y cierre de sesión.

### ***3.6.9 Flujo de interacción del especialista***

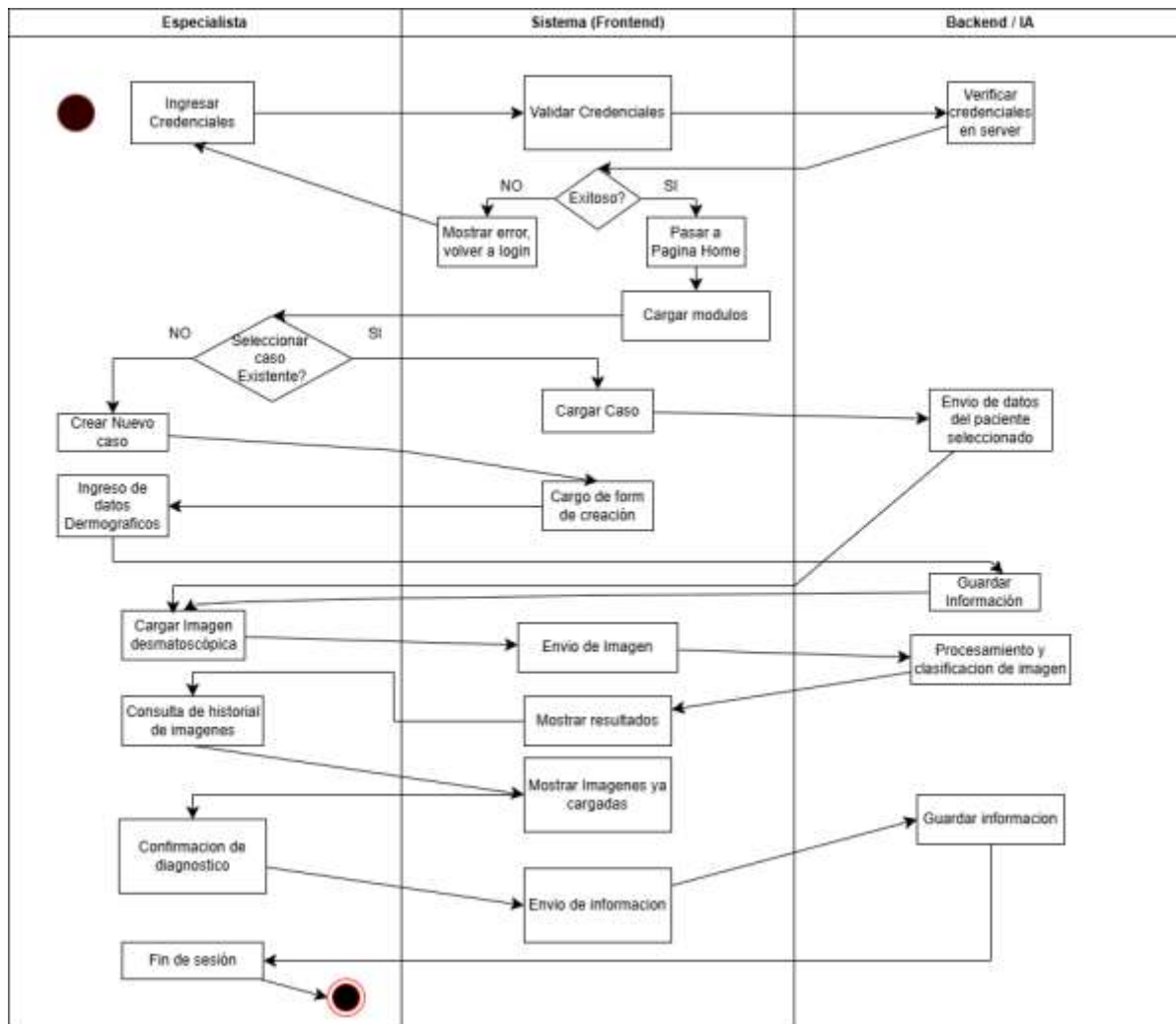
El flujo de uso principal de la plataforma se diseñó para minimizar el número de interacciones necesarias desde la autenticación hasta la obtención del resultado diagnóstico, en línea con el requerimiento RNF02. El especialista accede a la plataforma a través de la página de inicio de sesión; una vez autenticado, es dirigido al panel principal (HomePage), donde visualiza simultáneamente los tres módulos de trabajo: el historial de casos, la ficha del paciente y el panel de diagnóstico.

Al seleccionar un caso en el historial lateral, la información del paciente se propaga automáticamente al formulario central y al panel de diagnóstico a través del contexto compartido, sin necesidad de navegación adicional. Si se trata de un caso nuevo, el especialista puede crearlo desde el formulario central, completar los datos demográficos y proceder inmediatamente a la carga de la imagen dermatoscópica en el panel de diagnóstico.

El panel de diagnóstico permite al especialista cargar una imagen desde su dispositivo (PC o móvil), visualizarla en tamaño ampliado, consultar el historial de imágenes previas del caso con sus respectivas probabilidades asignadas por el modelo, y confirmar o rechazar la clasificación sugerida por la inteligencia artificial mediante un mecanismo de checkbox de validación. Tanto la clasificación como la probabilidad de malignidad son campos de solo lectura en la interfaz, dado que sus valores son determinados exclusivamente por el modelo y asignados desde el backend al momento del análisis.

En la Figura 22 se ve plasmado este flujo en el diagrama de actividades orientado a la aplicación.

**Figura 22.**  
*Diagrama de Actividades del Flujo de la Aplicación*



*Nota.* Diagrama de actividades del usuario en interacción de las demás partes del sistema

### ***3.6.10 Responsividad y diseño para dispositivos móviles***

El diseño responsivo de la interfaz se implementó mediante un sistema de cuadrícula de 12 columnas de Tailwind CSS con breakpoints condicionados por el prefijo lg (1024 píxeles). En dispositivos con resolución inferior a este umbral, los tres paneles del módulo principal se apilan verticalmente en el orden definido por los requerimientos clínicos: primero el historial de casos, seguido de la ficha del paciente y finalmente el panel de diagnóstico. En pantallas de escritorio, los paneles se distribuyen horizontalmente en proporciones de 3-6-3 columnas.

Adicionalmente, el mecanismo de carga de imágenes mediante el elemento HTML input de tipo file con el atributo `accept="image/*"` permite que, en dispositivos móviles, el sistema operativo nativo presente al usuario las opciones de galería, cámara o explorador de archivos sin requerir implementación adicional en el frontend, facilitando el uso clínico desde dispositivos móviles.

### ***3.6.11 Metodología de desarrollo***

El desarrollo del prototipo web se llevó a cabo mediante una metodología ágil iterativa e incremental, adaptada a las características y restricciones del proyecto de grado. Esta elección responde a la naturaleza exploratoria de la investigación, en la que los requisitos del sistema evolucionan a medida que se consolidan los resultados del modelo de inteligencia artificial, y a la necesidad de validar incrementalmente las funcionalidades con usuarios potenciales antes de la entrega final.

El ciclo de desarrollo se estructuró en torno al cronograma de actividades establecido en el plan de trabajo de grado (véase sección de cronograma), que distribuyó las tareas en 16 semanas

organizadas en seis fases secuenciales: preparación del conjunto de datos (semanas 1–3), preprocesamiento de imágenes (semanas 3–4), modelado y entrenamiento del modelo híbrido (semanas 4–10), evaluación del desempeño del modelo (semanas 9–11), desarrollo del prototipo web (semanas 11–15) y validación y documentación (semanas 15–16).

Dentro de la fase de desarrollo del prototipo web, el trabajo se organizó en iteraciones semanales con entregas parciales revisables. Cada iteración partió de la definición de un conjunto acotado de funcionalidades a implementar, derivadas de los requerimientos funcionales priorizados, y culminó con una revisión interna del estado del prototipo por parte del equipo de desarrollo y la directora del proyecto.

La primera iteración abordó la configuración del entorno de desarrollo (React + Vite + Tailwind CSS), la implementación del sistema de enrutamiento y la definición de los contextos globales de autenticación y gestión de casos. La segunda iteración se centró en los módulos de autenticación (inicio de sesión y registro) y en el diseño del panel principal con los tres componentes de trabajo. Las iteraciones posteriores completaron progresivamente la ficha del paciente, el panel de diagnóstico con soporte de imágenes y el módulo de perfil de usuario, incorporando en cada ciclo las retroalimentaciones derivadas de la revisión anterior.

La comunicación entre el equipo de desarrollo (tres estudiantes) se realizó de forma periódica, con sesiones de revisión que permitieron ajustar el alcance funcional del prototipo en función del avance del modelo de inteligencia artificial. Esta coordinación resultó especialmente relevante para definir los contratos de la API de predicción y los formatos de respuesta del modelo, que condicionan directamente el diseño del panel de diagnóstico.

### 3.6.12 Pruebas técnicas y de usabilidad

La validación del prototipo web se estructuró en dos dimensiones complementarias: pruebas técnicas orientadas a verificar el correcto funcionamiento de los componentes del sistema, y pruebas de usabilidad destinadas a evaluar la experiencia de uso desde la perspectiva del especialista médico.

**3.6.12.1 Pruebas técnicas.** Las pruebas técnicas se diseñaron para verificar el cumplimiento de los requerimientos funcionales y no funcionales del sistema. Se organizaron en tres categorías: pruebas de integración entre el frontend y el backend, pruebas de rendimiento del endpoint de predicción y pruebas de flujo de usuario en los principales casos de uso.

**Tabla 16. Pruebas técnicas del prototipo web**

ID	Categoría	Objetivo	RF/ RNF	Procedimiento	Resultado esperado
PT01	Autenticación	Verificar que el sistema rechaza credenciales incorrectas y otorga acceso con credenciales válidas.	RF01 / RNF01	Intentar inicio de sesión con credenciales válidas e inválidas.	Acceso concedido con credenciales válidas; mensaje de error con credenciales inválidas.
PT02	Protección de rutas	Verificar que las rutas /home y /perfil no son accesibles sin sesión activa.	RF01 / RNF01	Intentar acceder a /home sin autenticación.	El sistema muestra la pantalla de acceso restringido.
PT03	CRUD de casos	Verificar la creación, consulta, edición y eliminación de casos clínicos.	RF02	Ejecutar secuencialmente las operaciones de creación, lectura, modificación y	Todas las operaciones se reflejan correctamente en el historial y en la base de datos.

ID	Categoría	Objetivo	RF/ RNF	Procedimiento	Resultado esperado
				borrado sobre un caso de prueba.	
PT04	Carga de imagen	Verificar que la imagen se almacena correctamente y se muestra en el visor.	RF03 / RF05	Cargar una imagen desde PC y desde dispositivo móvil.	La imagen aparece en el visor principal y como miniatura en el historial fotográfico.
PT05	Predicción del modelo	Verificar que el endpoint de predicción retorna clasificación y probabilidad dentro del tiempo máximo permitido.	RF03 / RF04 / RNF03	Enviar una imagen al endpoint /predict y medir el tiempo de respuesta.	Respuesta recibida en menos de 5 segundos con clasificación y probabilidad válidas.
PT06	Confirmación de diagnóstico	Verificar que el campo de confirmación del médico se guarda correctamente.	RF07	Activar el checkbox de confirmación, guardar y recargar el caso.	El estado de confirmación persiste tras el guardado.
PT07	Responsividad	Verificar la visualización correcta en dispositivos móviles y de escritorio.	RNF02	Cargar la aplicación en resoluciones de 375px (móvil), 768px (tableta) y 1280px (escritorio).	En móvil, los paneles se apilan verticalmente; en escritorio, se distribuyen en tres columnas.
PT08	Fallo del modelo	Verificar el comportamiento cuando el modelo no está disponible.	RNF05	Deshabilitar el servicio de predicción e intentar cargar una imagen.	El sistema muestra un mensaje de error claro al usuario sin colapsar la aplicación.

**3.6.12.2 Pruebas de usabilidad.** Las pruebas de usabilidad se diseñaron para evaluar la experiencia del especialista médico al interactuar con el prototipo en condiciones representativas del entorno clínico. El objetivo principal fue identificar fricciones en el flujo de trabajo, ambigüedades en la interpretación de los resultados diagnósticos y oportunidades de mejora en la organización de la interfaz.

La metodología adoptada fue la de prueba de usabilidad con tareas guiadas, en la que cada participante completó un conjunto de tareas predefinidas mientras el equipo de investigación registraba el tiempo de realización, los errores cometidos y las expresiones verbales espontáneas. Al finalizar las tareas, se aplicó un cuestionario de satisfacción basado en la escala de usabilidad del sistema (SUS, por sus siglas en inglés), que permite cuantificar la percepción global de usabilidad en una escala de 0 a 100.

Las tareas definidas para la sesión de pruebas cubrieron los principales casos de uso del sistema: registro e inicio de sesión (CU01, CU02), creación de un nuevo caso clínico con datos demográficos (CU03), carga de una imagen dermatoscópica y visualización del resultado del modelo (CU06, CU07), confirmación del diagnóstico sugerido (CU08) y acceso al perfil de usuario para actualizar sus datos (CU10).

La evaluación de usabilidad fue realizada con el acompañamiento de la codirectora del proyecto, quien participó en el proceso de validación del sistema, aportando su criterio en relación con la curva de aprendizaje y la adecuación de la plataforma al contexto clínico. Como soporte de este proceso, se anexa la URL del formulario utilizado para la recolección de la información correspondiente a la encuesta aplicada.

<https://forms.gle/j7LhPTAAX8gBdqHW6>

### ***3.6.13 Evaluación y retroalimentación***

La evaluación del prototipo web integra los resultados de las pruebas técnicas y de usabilidad descritas en la sección anterior, con el objetivo de determinar en qué medida el sistema cumple con los requerimientos establecidos y satisface las necesidades del usuario clínico.

Desde la perspectiva técnica, los resultados de las pruebas se analizaron en relación con los criterios de aceptación definidos para cada requerimiento no funcional. El tiempo de respuesta del endpoint de predicción, la correcta visualización en dispositivos de distintas resoluciones, la integridad de los datos tras operaciones CRUD y el comportamiento del sistema ante fallos del modelo son los indicadores primarios de la evaluación técnica.

Desde la perspectiva de usabilidad, los resultados del cuestionario SUS se interpretan según los rangos establecidos en la literatura: una puntuación superior a 85 se considera excelente; entre 71 y 85, buena; entre 52 y 70, aceptable; e inferior a 52, deficiente. Adicionalmente, el análisis cualitativo de las observaciones registradas durante las sesiones de prueba permite identificar patrones recurrentes de dificultad y generar recomendaciones concretas de mejora para versiones posteriores del sistema.

La retroalimentación recopilada durante las pruebas de usabilidad fue categorizada en tres dimensiones: claridad de la interfaz (facilidad para interpretar los resultados del modelo y navegar entre módulos), eficiencia del flujo de trabajo (número de interacciones necesarias para completar una tarea) y confianza en el sistema (percepción del especialista sobre la fiabilidad de las predicciones del modelo y la seguridad de los datos del paciente).

Los hallazgos de esta evaluación se incorporan como insumo para la sección de conclusiones del trabajo de grado y como base para la definición de trabajos futuros, en particular

la integración completa del modelo de IA en el backend, la implementación del módulo de almacenamiento de imágenes en Supabase y la realización de pruebas de usabilidad con una muestra más amplia de especialistas en contextos clínicos reales.

### 3.7. Discusión de resultados

El modelo híbrido CNN+ViT alcanzó un AUC-ROC de 0,9731 [IC 95%: 0,9681–0,9777] sobre el conjunto de prueba ( $n = 3.800$ ), superando tanto al EfficientNet-B0 (AUC = 0,9628) como al DeiT-B/16 (AUC = 0,9700). Este resultado confirma la hipótesis central de la arquitectura: la fusión de representaciones locales (CNN) y globales (ViT) mediante una compuerta de atención aprendida captura información complementaria que ningún backbone puede extraer de forma aislada. Esta tendencia es coherente con lo reportado en la literatura.[5].

Como se expuso anteriormente, la métrica de mayor consecuencia es la sensibilidad o recall, en esta métrica el modelo híbrido alcanzó un valor de 0,9101 (91,01%) equivalente a 124 falsos negativos sobre 1.401 casos malignos del conjunto de prueba (8,9%). Esta cifra representa una reducción de 43 casos respecto a la CNN (167 FN; 11,9%) y de 28 respecto al ViT DeiT (152 FN; 10,8%).

El análisis de los pesos de atención muestra valores medios cercanos a 0,5 en ambas ramas, indicando que la compuerta opera en la región lineal de la sigmoide y actúa como modulador, no como selector. Sin embargo, el diferencial de clases es mayor en la rama ViT que en la CNN, lo que evidencia mayor capacidad discriminativa global, consistente con su mejor AUC individual (0,9700 vs. 0,9628). Los resultados de ablación confirman esto, al eliminar ViT se reduce más el rendimiento (-1,43 pp) que eliminando CNN (-1,27 pp), aunque ambas contribuyen al desempeño

final. Ambas caídas superan el umbral de significancia práctica de 1 pp habitualmente considerado en clasificación médica, confirmando que las dos ramas aportan información relevante.

El diseño del prototipo web busca trasladar el modelo al entorno clínico de forma accesible y segura. La arquitectura cliente-servidor desacoplada permite actualizar el motor de inferencia sin afectar la interfaz, mientras que el mecanismo fail-fast asegura que el servicio no opere si los pesos no se cargan correctamente. Estas decisiones de ingeniería están alineadas con las recomendaciones de traducción clínica en machine learning descritas en la literatura [29], que enfatizan la importancia de la robustez operacional y la trazabilidad de resultados en entornos médicos. El registro de diagnósticos con probabilidad de malignidad, imagen asociada y confirmación del especialista constituye además un punto de partida para la recolección prospectiva de datos locales, que podría alimentar futuras rondas de ajuste fino del modelo.

### ***3.7.1 Resultados de Funcionalidad de la Aplicación***

Los resultados de las pruebas funcionales evidencian un desempeño sólido del sistema en la mayoría de los escenarios evaluados. Las pruebas relacionadas con autenticación, protección de rutas, operaciones CRUD de casos clínicos, carga de imágenes, predicción del modelo, confirmación de diagnóstico y responsividad (PT01 a PT07) fueron satisfactorias, cumpliendo con los requerimientos funcionales y no funcionales definidos. Esto confirma que el sistema garantiza un acceso seguro, una correcta gestión de la información clínica y una adecuada visualización en distintos dispositivos.

No obstante, la prueba PT08, correspondiente al manejo de fallos del modelo, no fue superada, lo que indica una debilidad en la gestión de errores cuando el servicio de predicción no está

disponible. Este resultado sugiere la necesidad de implementar mecanismos de manejo de excepciones y retroalimentación al usuario que permitan mantener la estabilidad y la experiencia de uso ante este tipo de situaciones.

**Tabla 17. Resultados pruebas técnicas del prototipo web**

<b>ID</b>	<b>Categoría</b>	<b>Objetivo</b>	<b>RF/ RNF</b>	<b>Procedimiento</b>	<b>Resultado Obtenido</b>
PT01	Autenticación	Verificar que el sistema rechaza credenciales incorrectas y otorga acceso con credenciales válidas.	RF01 / RNF01	Intentar inicio de sesión con credenciales válidas e inválidas.	Válido
PT02	Protección de rutas	Verificar que las rutas /home y /perfil no son accesibles sin sesión activa.	RF01 / RNF01	Intentar acceder a /home sin autenticación.	Válido
PT03	CRUD de casos	Verificar la creación, consulta, edición y eliminación de casos clínicos.	RF02	Ejecutar secuencialmente las operaciones de creación, lectura, modificación y borrado sobre un caso de prueba.	Válido
PT04	Carga de imagen	Verificar que la imagen se almacena correctamente y se muestra en el visor.	RF03 / RF05	Cargar una imagen desde PC y desde dispositivo móvil.	Válido
PT05	Predicción del modelo	Verificar que el endpoint de predicción retorna clasificación y probabilidad dentro del tiempo máximo permitido.	RF03 / RF04 / RNF03	Enviar una imagen al endpoint /predict y medir el tiempo de respuesta.	Válido

ID	Categoría	Objetivo	RF/ RNF	Procedimiento	Resultado Obtenido
PT06	Confirmación de diagnóstico	Verificar que el campo de confirmación del médico se guarda correctamente.	RF07	Activar el checkbox de confirmación, guardar y recargar el caso.	Válido
PT07	Responsividad	Verificar la visualización correcta en dispositivos móviles y de escritorio.	RNF02	Cargar la aplicación en resoluciones de 375px (móvil), 768px (tableta) y 1280px (escritorio).	Válido
PT08	Fallo del modelo	Verificar el comportamiento cuando el modelo no está disponible.	RNF05	Deshabilitar el servicio de predicción e intentar cargar una imagen.	No Válido

#### 4. Conclusiones

En conclusión, la fusión de representaciones locales mediante una red convolucional y representaciones globales por medio de un vision transformer realizada a partir de una compuerta de atención aprendida —gate attention— produjo un modelo que supera sistemáticamente a ambos backbones individuales: AUC-ROC de 0,9731 (Híbrido) vs 0,9628 (CNN) vs 0,9700 (DeiT-B/16). El estudio de ablación cuantificó que la contribución de cada rama supera el umbral de significancia práctica de 1 pp en clasificación médica (−1,43 pp sin ViT; −1,27 pp sin CNN), confirmando que las dos fuentes de información son complementarias y no redundantes.

En términos clínicos, el modelo híbrido logró identificar correctamente el 91,01% de las 1.401 lesiones malignas del conjunto de prueba (124 falsos negativos). Esto se traduce en una

capacidad de detección que se ubica por encima del rango típico reportado para dermatólogos sin asistencia computacional [4], aunque la validación de esta comparación en condiciones clínicas reales sigue siendo necesaria.

La calibración probabilística del modelo híbrido con valor de 0,031 es excelente según el umbral estándar de la literatura, esto permite utilizar la probabilidad continua de malignidad como guía para decidir qué tan urgente es derivar al paciente, una propiedad clave para integrarlo en el trabajo diario clínico real.

Por último, el prototipo web desarrollado sobre una arquitectura moderna satisface los requerimientos funcionales definidos —incluyendo la gestión estructurada de pacientes, el análisis automatizado de imágenes mediante IA, el almacenamiento longitudinal de historial fotográfico y la confirmación clínica del diagnóstico— e incorpora el modelo híbrido como un servicio de inferencia independiente, optimizado para tiempos de respuesta reducidos y procesamiento eficiente en tiempo real. Esta separación entre capa de presentación y lógica de inferencia favorece la escalabilidad, mantenibilidad y actualización independiente de los componentes del sistema.

El mecanismo de colaboración humano-IA sigue el marco de Tschandl [23], donde el especialista valida la clasificación sugerida por el algoritmo, manteniendo el criterio médico como instancia final de decisión y aprovechando la capacidad del modelo para mejorar la consistencia diagnóstica, reducir la variabilidad interobservador y optimizar el proceso de priorización clínica.

## **5. Trabajo futuro**

En primer lugar, la validación externa con datos locales y adaptación de dominio se considera crítica, dado que el modelo ha sido entrenado sobre el conjunto ISIC 2019, el cual

presenta un sesgo geográfico respecto a la población objetivo. Es necesario construir o integrar un corpus dermatoscópico local que permita cuantificar el dataset shift y evaluar la degradación del desempeño, así como explorar estrategias que preserven simultáneamente la capacidad discriminativa y la calibración probabilística del modelo en el nuevo dominio.

En segundo lugar, la validación clínica prospectiva con especialistas constituye un paso indispensable para la transición del sistema a un entorno real de atención sanitaria. Este proceso debe orientarse a la evaluación integral del sistema en condiciones operativas, incluyendo métricas de desempeño clínico e integración en el flujo de trabajo del dermatólogo.

Finalmente, la integración con sistemas de información hospitalaria es fundamental para su adopción efectiva en entornos clínicos. Esta integración permitiría incorporar las predicciones del modelo dentro del historial médico electrónico del paciente, garantizando trazabilidad, continuidad asistencial y alineación con los flujos de trabajo establecidos en instituciones de salud. Asimismo, habilita la escalabilidad del sistema hacia escenarios de consultas remotas y atención distribuida.

### Referencias Bibliográficas

- [1] World Cancer Research Fund, "Skin cancer statistics," World Cancer Research Fund, Jul. 28, 2025. <https://www.wcrf.org/preventing-cancer/cancer-statistics/skin-cancer-statistics/>
- [2] P. Karp, K. Karp, M. Kaździela, R. Zajdel, and A. Żebrowska, "The importance of early detection and prevention of atypical skin lesions and other melanoma risk factors in a younger population," *Cancers*, vol. 16, no. 24, p. 4264, 2024, doi:10.3390/cancers16244264.
- [3] A. Lallas and G. Argenziano, "Artificial intelligence and melanoma diagnosis: Ignoring human nature may lead to false predictions," *Dermatology Practical & Conceptual*, pp. 249–251, 2018, doi:10.5826/dpc.0804a01.
- [4] T. J. Brinker 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, pp. 47–54, 2019, doi:10.1016/j.ejca.2019.04.001.
- [5] O. Akinrinade and C. Du, "Skin cancer detection using deep machine learning techniques," *Intelligence-Based Medicine*, vol. 11, p. 100191, 2025, doi:10.1016/j.ibmed.2024.100191.
- [6] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, 2018, doi:10.1038/sdata.2018.161.
- [7] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020. [Online]. Available: <http://arxiv.org/abs/2010.11929>
- [8] M. A. Khan, M. Y. Javed, A. Sharif, T. Saba, M. T. Iqbal, and J. M. R. Tavares, "Skin lesion segmentation and classification: A unified framework of deep neural networks for skin

- cancer diagnosis," *Healthcare*, vol. 10, no. 7, p. 1183, 2022, doi:10.3390/healthcare10071183.
- [9] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," arXiv [cs.CV], 2021.
- [10] Z. Shao et al., "TransMIL: Transformer based correlated multiple instance learning for whole slide image classification," arXiv [cs.CV], 2021.
- [11] Y. Bazi, L. Bashmal, M. M. Rahhal, R. A. Dayil, and N. Al Ajlan, "Vision Transformers for Remote Sensing Scene Classification," *Remote Sensing*, vol. 13, no. 3, p. 516, 2021.
- [12] A. Lomas, J. Leonardi-Bee, and F. Bath-Hextall, "A systematic review of worldwide incidence of nonmelanoma skin cancer," *British Journal of Dermatology*, vol. 166, no. 5, pp. 1069–1080, Jan. 2012, doi:10.1111/j.1365-2133.2012.10830.x.
- [13] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, Feb. 2021, doi:10.3322/caac.21660.
- [14] Liga Colombiana contra el Cáncer, "Datos sobre cáncer de piel en Colombia," Liga Contra el Cáncer, 2025. <https://www.ligacancercolombia.org/educacion/datos-cancer-de-piel/>
- [15] C. A. Bonilla et al., "Incidencia de cáncer de piel en Colombia, años 2003–2007," *Piel: Formación Continuada en Dermatología*, vol. 26, no. 8, pp. 429–436, 2011. doi:10.1016/j.piel.2011.07.001.
- [16] Instituto Nacional de Cancerología, *Manual para la detección temprana del cáncer de piel en Colombia*, MinSalud, Bogotá, 2012. [Online]. Available: [https://www.cancer.gov.co/recursos\\_user/files/libros/archivos/Piel](https://www.cancer.gov.co/recursos_user/files/libros/archivos/Piel)

- [17] Mayo Clinic, "Cáncer de piel – Síntomas y causas," Mayo Clinic, 2025. <https://www.mayoclinic.org/es/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605>
- [18] American Cancer Society, "¿Qué son los cánceres de piel de células basales y de células escamosas?," 2019. <https://www.cancer.org/es/cancer/tipos/cancer-de-piel-de-celulas-basales-y-escamosas/>
- [19] E. Kalloniati, D. Cavouras, K. Plachouri, E. Geropoulou, G. Sakellaropoulos, and S. Georgiou, "Clinical, dermoscopic and histological assessment of melanocytic lesions: a comparative study of the accuracy of the diagnostic methods," Dec. 01, 2021. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9894303/>
- [20] J. Dinnes et al., "Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults," *Cochrane Database of Systematic Reviews*, no. 12, CD011902, 2018. doi:10.1002/14651858.CD011902.pub2
- [21] V. Nikolaou et al., "Clinical, dermoscopic and histological assessment of melanocytic lesions in adults: Diagnostic accuracy study," *Dermatology Practical & Conceptual*, vol. 13, no. 1, e2023041, 2023. <https://pubmed.ncbi.nlm.nih.gov/36743868/>
- [22] S. Haggenmüller et al., "Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts," *European Journal of Cancer*, vol. 156, pp. 202–216, 2021.
- [23] P. Tschandl, C. Rinner, Z. Apalla et al., "Human-computer collaboration for skin cancer recognition," *Nature Medicine*, vol. 26, pp. 1229–1234, 2020. doi:10.1038/s41591-020-0942-0
- [24] Policlínico San Miguel, "Cáncer de piel," Policlínico San Miguel, s. f. <https://policlinicosanmiguel.com/especialidades-medicas/dermatologia/cancer-de-piel/>

- [25] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, 2015.
- [26] I. Chaves, M. Talo, and P. F. Jaeger, "Vision transformers in medical computer vision—A survey," arXiv preprint arXiv:2301.03505, 2023.
- [27] S. Almotairi, E. Badr, I. Abdelbaky, M. Elhakeem, and M. Abdul Salam, "Hybrid transformer-CNN model for accurate prediction of peptide hemolytic potential," *Scientific Reports*, vol. 14, p. 14263, 2024. doi:10.1038/s41598-024-63446-5
- [28] World Health Organization, *Global strategy on digital health 2020–2025*, WHO, 2020. <https://www.who.int/docs/default-source/documents/gS4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf>
- [29] M. Poddar et al., "An operational guide to translational clinical machine learning in academic medical centers," *npj Digital Medicine*, 2024. <https://www.nature.com/articles/s41746-024-01094-9>
- [30] T. A. D. Graham et al., "Effects of a web-based patient portal on health outcomes and patient engagement," *Journal of Medical Internet Research*, 2020. <https://www.jmir.org/2020/5/e17955/>
- [31] Ministerio de Salud y Protección Social (Colombia), *Nuevo marco reglamentario para la telesalud en Colombia (Resolución 2654 de 2019)*, 2019. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/OT/nuevo-marco-reglamentario-para-la-telesalud-en-colombia-18122019.pdf>
- [32] A. Takiddin et al., "Artificial intelligence for skin cancer detection: A scoping review," *Journal of Medical Internet Research*, 2021. <https://www.jmir.org/2021/11/e22934/>

- [33] American Cancer Society, Survival Rates for Melanoma Skin Cancer by Stage, 2023.
- [34] Perez, L., & Wang, J. “The effectiveness of data augmentation in image classification using deep learning”. arXiv preprint arXiv:1712.04621, 2017. <https://arxiv.org/abs/1712.04621>
- [35] Efron, B., & Tibshirani, R. J. An introduction to the bootstrap, Chapman & Hall/CRC, 2023.
- [36] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. Proceedings of the 34th International Conference on Machine Learning (ICML), 70, 1321–1330. <https://arxiv.org/abs/1706.04599>
- [37] Wightman, R. PyTorch Image Models (timm). GitHub, 2017. <https://github.com/rwightman/pytorch-image-models>
- [38] Goodfellow, I., Bengio, Y., & Courville, A. Deep learning. MIT Press, 2016. <https://www.deeplearningbook.org>
- [39] Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale”. International Conference on Learning Representations, 2021. <https://arxiv.org/abs/2010.11929>
- [40] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. Training data-efficient image transformers & distillation through attention., 139, 10347–10357. 2021. <https://arxiv.org/abs/2012.12877>
- [41] Chen, C.-F., Fan, Q., & Panda, R. CrossViT: Cross-attention multi-scale vision transformer for image classification, 357–366; 2021. <https://arxiv.org/abs/2103.14899>
- [42] Snoek, J., Larochelle, H., & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. Advances in Neural Information Processing Systems, 25, 2951–2959. 2012. <https://arxiv.org/abs/1206.2944>



## Apéndices

### APÉNDICE A. TABLAS DE REQUERIMIENTOS FUNCIONALES Y NO FUNCIONALES Y CASOS DE USO

#### *Requerimiento Funcional 1*

<i>Identificación del requerimiento</i>	RF01
<i>Nombre del requerimiento</i>	Gestión de autenticación
<i>Descripción del requerimiento</i>	El sistema debe permitir a los especialistas iniciar sesión y cerrar sesión de forma segura para proteger los datos de los pacientes.
<i>Prioridad del requerimiento</i>	Alta

#### *Requerimiento Funcional 2*

<i>Identificación del requerimiento</i>	RF02
<i>Nombre del requerimiento</i>	Gestión de Pacientes
<i>Descripción del requerimiento</i>	El sistema debe permitir el registro, consulta, actualización y eliminación de la información demográfica de los pacientes.
<i>Prioridad del requerimiento</i>	Media

#### *Requerimiento Funcional 3*

<i>Identificación del requerimiento</i>	RF03
<i>Nombre del requerimiento</i>	Análisis de Imágenes por IA

<i>Descripción del requerimiento</i>	El sistema debe permitir la carga de imágenes dermatoscópicas y procesarlas mediante el modelo híbrido para obtener una predicción automática.
<i>Prioridad del requerimiento</i>	Alta

**Requerimiento Funcional 4**

<i>Identificación del requerimiento</i>	RF04
<i>Nombre del requerimiento</i>	Visualización de Resultados
<i>Descripción del requerimiento</i>	El sistema debe mostrar la clasificación (Maligno/Benigno) y el porcentaje de probabilidad obtenido por el modelo.
<i>Prioridad del requerimiento</i>	Alta

**Requerimiento Funcional 5**

<i>Identificación del requerimiento</i>	RF05
<i>Nombre del requerimiento</i>	Historial Clínico Fotográfico
<i>Descripción del requerimiento</i>	El sistema debe permitir almacenar y visualizar un historial de múltiples imágenes y diagnósticos por cada paciente.
<i>Prioridad del requerimiento</i>	Alta

**Requerimiento Funcional 6**

<i>Identificación del requerimiento</i>	RF06
<i>Nombre del requerimiento</i>	Gestión de Imágenes
<i>Descripción del requerimiento</i>	El sistema debe permitir la eliminación de imágenes específicas del historial si el médico lo considera necesario.
<i>Prioridad del requerimiento</i>	Media

**Requerimiento Funcional 7**

<i>Identificación del requerimiento</i>	RF07
<i>Nombre del requerimiento</i>	Confirmación de diagnóstico
<i>Descripción del requerimiento</i>	El sistema debe permitir al especialista confirmar o rechazar la clasificación sugerida por el modelo de IA, registrando dicha validación como parte del historial del caso.
<i>Prioridad del requerimiento</i>	Media

**Requerimiento Funcional 8**

<i>Identificación del requerimiento</i>	RF08
<i>Nombre del requerimiento</i>	Gestión del perfil de usuario
<i>Descripción del requerimiento</i>	El sistema debe permitir al especialista visualizar y actualizar la información de su cuenta (nombre y correo electrónico), así como cerrar sesión de forma segura desde la interfaz de perfil.
<i>Prioridad del requerimiento</i>	Baja

**Requerimiento No Funcional 1**

<i>Identificación del requerimiento</i>	RNF01
<i>Nombre del requerimiento</i>	Seguridad de Datos
<i>Descripción del requerimiento</i>	Toda la información sensible y las imágenes deben almacenarse de forma cifrada en Supabase siguiendo estándares de privacidad de datos médicos.
<i>Prioridad del requerimiento</i>	Alta

**Requerimiento No Funcional 2**

<i>Identificación del requerimiento</i>	RNF02
<i>Nombre del requerimiento</i>	Usabilidad
<i>Descripción del requerimiento</i>	La interfaz debe ser intuitiva y minimalista, permitiendo realizar un diagnóstico en menos de 3 clics tras la creación del paciente.
<i>Prioridad del requerimiento</i>	Alta

**Requerimiento No Funcional 3**

<i>Identificación del requerimiento</i>	RNF03
<i>Nombre del requerimiento</i>	Rendimiento

<i>Descripción del requerimiento</i>	El tiempo de respuesta de la predicción de la IA no debe superar los 5 segundos (dependiendo de la latencia de red).
<i>Prioridad del requerimiento</i>	Alta

***Requerimiento No Funcional 4***

<i>Identificación del requerimiento</i>	RNF04
<i>Nombre del requerimiento</i>	Disponibilidad
<i>Descripción del requerimiento</i>	El backend debe estar alojado en una infraestructura que garantice que el servicio sea accesible durante la jornada laboral del especialista.
<i>Prioridad del requerimiento</i>	Alta

***Requerimiento No Funcional 5***

<i>Identificación del requerimiento</i>	RNF05
<i>Nombre del requerimiento</i>	Integridad
<i>Descripción del requerimiento</i>	El sistema debe asegurar que los pesos del modelo híbrido sean cargados correctamente al inicio del servidor; de lo contrario, el servicio de predicción no debe estar disponible.
<i>Prioridad del requerimiento</i>	Alta

**Requerimiento No Funcional 6**

<i>Identificación del requerimiento</i>	RNF06
<i>Nombre del requerimiento</i>	Mantenibilidad
<i>Descripción del requerimiento</i>	El código fuente del frontend y del backend debe estar organizado en módulos cohesivos y documentados, facilitando la incorporación de futuras mejoras tales como nuevos modelos de clasificación o la integración con sistemas de información hospitalaria.
<i>Prioridad del requerimiento</i>	Media

*Caso de Uso 1*

<b>Identificador</b>	CU01
<b>Nombre</b>	Iniciar sesión
<b>Descripción</b>	El especialista ingresa su correo electrónico y contraseña en la página de inicio de sesión y accede al panel principal del sistema.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	El usuario no está autenticado.
<b>Flujo principal</b>	1. El especialista accede a la página /inicio.2. Ingresar sus credenciales (correo y contraseña).3. El sistema valida las credenciales contra la base de datos.4. Si son correctas, genera un token JWT y redirige al panel principal (/home).5. Si son incorrectas, muestra un mensaje de error descriptivo.
<b>Postcondición</b>	El especialista queda autenticado y accede al panel principal.
<b>Flujo alternativo</b>	Credenciales incorrectas: el sistema muestra el mensaje 'Correo o contraseña incorrectos' sin revelar cuál de los dos campos es erróneo.
<b>RF asociado</b>	RF01

*Caso de Uso 2*

<b>Identificador</b>	CU02
<b>Nombre</b>	Registrar cuenta
<b>Descripción</b>	Un nuevo especialista crea su cuenta en el sistema proporcionando nombre, correo electrónico y contraseña.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	El usuario no posee cuenta registrada.
<b>Flujo principal</b>	1. El especialista accede a la página /registro.2. Completa el formulario con nombre, correo, contraseña y confirmación de contraseña.3. El sistema valida que el correo no esté registrado previamente y que las contraseñas coincidan.4. Si la validación es exitosa, crea el registro y redirige al panel principal.
<b>Postcondición</b>	La cuenta queda creada y el especialista accede autenticado al sistema.
<b>Flujo alternativo</b>	Correo ya registrado: el sistema informa que ya existe una cuenta asociada a ese correo.
<b>RF asociado</b>	RF01

*Caso de Uso 3*

<b>Identificador</b>	CU03
<b>Nombre</b>	Registrar nuevo caso clínico
<b>Descripción</b>	El especialista crea un nuevo registro de paciente con sus datos demográficos y clínicos básicos.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	El especialista está autenticado. No hay ningún caso seleccionado en el historial.
<b>Flujo principal</b>	1. El especialista hace clic en 'Crear nuevo caso' en el formulario central.2. Completa los campos: nombre, apellido, fecha de nacimiento, diagnóstico, estado y notas.3. Hace clic en 'Guardar'.4. El sistema crea el registro y lo agrega al historial lateral.
<b>Postcondición</b>	El nuevo caso queda registrado y disponible en el historial de casos.
<b>Flujo alternativo</b>	Campos obligatorios vacíos (nombre o apellido): el sistema no procesa el guardado.
<b>RF asociado</b>	RF02

*Caso de Uso 4*

<b>Identificador</b>	CU04
<b>Nombre</b>	Consultar y editar caso clínico
<b>Descripción</b>	El especialista consulta los datos de un paciente existente y, si es necesario, los actualiza.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	El especialista está autenticado y hay casos registrados en el historial.
<b>Flujo principal</b>	1. El especialista selecciona un caso en el historial lateral.2. El sistema muestra los datos del paciente en modo lectura en el formulario central.3. El especialista hace clic en 'Editar'.4. Modifica los campos deseados y hace clic en 'Guardar'.5. El sistema actualiza el registro y vuelve al modo lectura.
<b>Postcondición</b>	Los datos del caso quedan actualizados en el sistema.
<b>Flujo alternativo</b>	El especialista puede cancelar la edición; los cambios no se persisten.
<b>RF asociado</b>	RF02

*Caso de Uso 5*

<b>Identificador</b>	CU05
<b>Nombre</b>	Eliminar caso clínico
<b>Descripción</b>	El especialista elimina un caso clínico del historial.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	El especialista está autenticado y hay un caso seleccionado.
<b>Flujo principal</b>	1. El especialista selecciona el caso en el historial.2. Hace clic en 'Borrar' en el formulario central.3. El sistema solicita confirmación mediante un segundo clic.4. Confirmada la acción, el caso es eliminado del historial.
<b>Postcondición</b>	El caso es eliminado y el panel de diagnóstico vuelve al estado vacío.
<b>Flujo alternativo</b>	Si el especialista no confirma la acción, el caso no se elimina.
<b>RF asociado</b>	RF02

*Caso de Uso 6*

<b>Identificador</b>	CU06
<b>Nombre</b>	Cargar imagen dermatoscópica
<b>Descripción</b>	El especialista carga una imagen de la lesión cutánea del paciente al panel de diagnóstico.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	Hay un caso seleccionado en el historial.
<b>Flujo principal</b>	1. El especialista hace clic en 'Subir imagen desde PC' en el panel de diagnóstico.2. El sistema abre el selector de archivos nativo del dispositivo.3. El especialista selecciona una imagen (formatos aceptados: JPG, PNG).4. El sistema almacena la imagen, la muestra en el visor principal y la agrega al historial de imágenes del caso.
<b>Postcondición</b>	La imagen queda asociada al caso y visible en el panel de diagnóstico.
<b>Flujo alternativo</b>	Si no hay caso seleccionado, el panel muestra el estado vacío sin opción de carga.
<b>RF asociado</b>	RF03, RF05

*Caso de Uso 7*

<b>Identificador</b>	CU07
<b>Nombre</b>	Obtener predicción del modelo de IA
<b>Descripción</b>	El sistema procesa la imagen cargada mediante el modelo híbrido y retorna la clasificación y probabilidad de malignidad.
<b>Actor</b>	Sistema (invocado automáticamente tras la carga de imagen)
<b>Precondición</b>	Se ha cargado una imagen dermatoscópica asociada a un caso.
<b>Flujo principal</b>	1. El frontend envía la imagen al endpoint de predicción del backend (FastAPI).2. El backend preprocesa la imagen y ejecuta la inferencia del modelo CNN-ViT.3. El modelo retorna la clasificación (Benigno/Maligno) y la probabilidad (0-100%).4. El backend almacena el resultado y lo retorna al frontend.5. El panel de diagnóstico actualiza la barra de probabilidad y el badge de clasificación.
<b>Postcondición</b>	La clasificación y probabilidad quedan registradas en el caso y visualizadas en el panel.
<b>Flujo alternativo</b>	Si el modelo no está disponible (RNF05), el sistema informa el error al usuario.
<b>RF asociado</b>	RF03, RF04

*Caso de Uso 8*

<b>Identificador</b>	CU08
<b>Nombre</b>	Confirmar diagnóstico
<b>Descripción</b>	El especialista valida la clasificación sugerida por el modelo de IA como diagnóstico confirmado del caso.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	El caso tiene una clasificación asignada por el modelo y el panel está en modo edición.
<b>Flujo principal</b>	1. El especialista entra al modo edición del panel de diagnóstico.2. Marca el checkbox 'Diagnóstico confirmado por el doctor'.3. Hace clic en 'Guardar'.4. El sistema registra la confirmación en el historial del caso.
<b>Postcondición</b>	El caso queda marcado con diagnóstico confirmado y el checkbox se visualiza en verde.
<b>Flujo alternativo</b>	El campo de confirmación es de solo lectura en modo visualización.
<b>RF asociado</b>	RF07

*Caso de Uso 9*

<b>Identificador</b>	CU09
<b>Nombre</b>	Visualizar historial fotográfico
<b>Descripción</b>	El especialista consulta las imágenes previas cargadas para un caso y navega entre ellas.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	El caso seleccionado tiene más de una imagen registrada.
<b>Flujo principal</b>	1. El especialista selecciona el caso en el historial.2. El panel de diagnóstico muestra la imagen más reciente en el visor principal.3. Las imágenes previas se presentan como miniaturas con su probabilidad asociada.4. El especialista hace clic en una miniatura para visualizarla en el visor principal.
<b>Postcondición</b>	El especialista puede navegar entre todas las imágenes históricas del caso.
<b>Flujo alternativo</b>	N/A
<b>RF asociado</b>	RF05

*Caso de Uso 10*

<b>Identificador</b>	CU10
<b>Nombre</b>	Gestionar perfil de usuario
<b>Descripción</b>	El especialista consulta y actualiza sus datos de perfil o cierra sesión.
<b>Actor</b>	Especialista médico
<b>Precondición</b>	El especialista está autenticado.
<b>Flujo principal</b>	1. El especialista hace clic en su nombre o avatar en la barra de navegación.2. El menú desplegable muestra opciones: 'Ver perfil', 'Editar perfil' y 'Cerrar sesión'.3. Al acceder al perfil, puede visualizar nombre, correo, ID y fecha de registro.4. En modo edición, puede modificar nombre y correo y guardar los cambios.
<b>Postcondición</b>	Los datos actualizados quedan reflejados en el estado de sesión y en la base de datos.
<b>Flujo alternativo</b>	ID y fecha de registro son campos de solo lectura.
<b>RF asociado</b>	RF08

## APÉNDICE B . MANUAL DE USUARIO – DERMASCAN

### B.1 Introducción

DermaScan es el nombre dado al prototipo del sistema que busca como objetivo ser una plataforma web de apoyo al diagnóstico de cáncer de piel diseñada para especialistas médicos. Permite gestionar casos clínicos de pacientes, cargar imágenes dermatoscópicas y obtener una clasificación automática de la lesión como benigna o maligna mediante un modelo híbrido de aprendizaje profundo (CNN–ViT), entrenado sobre el corpus ISIC 2019.

Este manual describe paso a paso las funcionalidades disponibles en la versión 1.0 del prototipo, orientando al usuario en el uso correcto de cada módulo. DermaScan es una herramienta de apoyo al criterio clínico y no reemplaza el juicio del especialista.

### B.2 Requisitos de acceso

Para utilizar el prototipo no se requiere la instalación de ningún software adicional. El sistema es accesible desde cualquier navegador web moderno con conexión a internet. Los requisitos mínimos se encuentran en la Tabla 18.

**Tabla 18. Requisitos de acceso a la plataforma**

Requisito	Especificación
Navegador web	Google Chrome 110+, Mozilla Firefox 110+ o Safari 16+
Conexión a internet	Velocidad mínima de 5 Mbps para carga fluida de imágenes
Dispositivo	PC, laptop, tableta o teléfono inteligente con pantalla táctil
Resolución mínima	360 × 640 píxeles (móvil) o 1024 × 768 píxeles (escritorio)
Cuenta de usuario	Correo electrónico válido y contraseña de mínimo 6 caracteres

*Nota.* La plataforma se adapta automáticamente a dispositivos móviles, apilando los paneles de forma vertical en pantallas pequeñas.

### B.3 Estructura general de la plataforma

La Tabla 19 muestra la organización de este sistema basado en cuatro áreas principales accesibles desde la interfaz.

**Tabla 19. Estructura general de la plataforma**

Área	Componente	Función principal
Historial de casos	Panel lateral izquierdo	Lista los últimos 15 casos del especialista con opción de selección.
Ficha del paciente	Panel central	Muestra y permite editar los datos demográficos y clínicos del caso seleccionado o crear uno nuevo.
Panel de diagnóstico	Panel lateral derecho	Gestiona las imágenes dermatoscópicas y presenta la clasificación y probabilidad del modelo de IA.
Perfil de usuario	Menú superior derecho	Permite consultar y editar los datos del especialista y cerrar sesión.

### B.4 Módulo de autenticación

#### B.4.1 Registro de cuenta nueva

Si es la primera vez que accede a DermaScan, debe crear una cuenta siguiendo estos pasos:

1. Acceda a la URL de la plataforma. El sistema le mostrará automáticamente la página de inicio de sesión.
2. Haga clic en el enlace Regístrese aquí, ubicado en la parte inferior del formulario.
3. Complete los campos: nombre completo, correo electrónico, contraseña (mínimo 6 caracteres) y confirmar contraseña.
4. Haga clic en el botón Crear cuenta.
5. Si el registro es exitoso, el sistema lo redirigirá automáticamente al panel principal.

**Nota.** El correo electrónico ingresado será su identificador único en el sistema. No podrá registrar dos cuentas con el mismo correo.

#### ***B.4.2 Inicio de sesión***

6. Acceda a la URL de la plataforma.
7. Ingrese su correo electrónico y contraseña en los campos correspondientes.
8. Haga clic en el botón Iniciar sesión.
9. Si las credenciales son correctas, será redirigido al panel principal (/home).

**Credenciales de demostración:** Para la sesión de prueba, puede utilizar el correo doctor@demo.com y la contraseña 123456.

#### ***B.4.3 Cierre de sesión***

10. Haga clic sobre su nombre o avatar en la esquina superior derecha de la barra de navegación.
11. En el menú desplegable, seleccione Cerrar sesión.
12. El sistema cerrará la sesión y lo redirigirá a la página de inicio de sesión.

### **B.5 Módulo de gestión de casos clínicos**

#### ***B.5.1 Visualización del historial de casos***

Al ingresar al panel principal, el historial lateral izquierdo muestra los últimos 15 casos registrados por el especialista. Cada tarjeta del historial presenta el nombre y apellido del paciente, el diagnóstico clínico registrado, el estado del caso (Activo, Seguimiento o Cerrado) y la fecha de nacimiento.

Para seleccionar un caso, haga clic sobre su tarjeta correspondiente. El sistema cargará automáticamente los datos del paciente en el formulario central y en el panel de diagnóstico. Para deseleccionar el caso activo, haga clic nuevamente sobre la misma tarjeta.

### ***B.5.2 Creación de un nuevo caso***

13. En el panel central, asegúrese de que no haya ningún caso seleccionado. El formulario mostrará el estado vacío con el botón + Crear nuevo caso.
14. Haga clic en + Crear nuevo caso.
15. Complete los campos del formulario: nombre, apellido, fecha de nacimiento, estado (Activo / Seguimiento / Cerrado), diagnóstico clínico y notas u observaciones adicionales.
16. Haga clic en el botón Guardar.
17. El nuevo caso aparecerá en la parte superior del historial lateral.

**Campos obligatorios:** Nombre y apellido. Los demás campos son opcionales pero se recomienda completarlos para un historial clínico completo.

### ***B.5.3 Edición de un caso existente***

18. Seleccione el caso en el historial lateral.
19. En el formulario central, haga clic en el botón Editar (ícono de lápiz).
20. Modifique los campos que desee actualizar.
21. Haga clic en Guardar para confirmar los cambios, o en Cancelar para descartarlos.

### ***B.5.4 Eliminación de un caso***

22. Seleccione el caso en el historial lateral.
23. En el formulario central, haga clic en el botón Borrar.

24. El sistema solicitará confirmación mostrando el botón ¿Confirmar?. Haga clic nuevamente para confirmar la eliminación.

25. Si desea cancelar la operación, haga clic en Cancelar antes de confirmar.

**Advertencia:** La eliminación de un caso es permanente e incluye todas las imágenes y resultados diagnósticos asociados. Esta acción no puede deshacerse.

## **B.6 Módulo de diagnóstico dermatológico**

El panel de diagnóstico, ubicado en la columna derecha del panel principal, permite gestionar las imágenes dermatoscópicas del paciente seleccionado y visualizar los resultados del modelo de inteligencia artificial. Este panel permanece en estado vacío mientras no haya un caso seleccionado.

### ***B.6.1 Carga de imagen dermatoscópica***

26. Seleccione el caso del paciente en el historial lateral.

27. En el panel de diagnóstico, haga clic en el botón Subir imagen desde PC.

28. El sistema abrirá el selector de archivos de su dispositivo. En dispositivos móviles, se presentarán las opciones de galería, cámara o explorador de archivos.

29. Seleccione la imagen en formato JPG o PNG.

30. La imagen se mostrará en el visor principal del panel y será procesada automáticamente por el modelo de IA.

**Recomendación clínica:** Para obtener los mejores resultados del modelo, utilice imágenes dermatoscópicas de alta resolución, bien enfocadas, con iluminación uniforme y sin obstrucciones sobre la lesión.

### ***B.6.2 Visualización del resultado diagnóstico***

Una vez procesada la imagen, el panel de diagnóstico mostrará dos indicadores de resultado generados por el modelo híbrido CNN-ViT:

1. Clasificación: indicador visual (badge de color) que muestra la categoría asignada a la lesión. Benigno se presenta en verde, Maligno en rojo y Sin clasificar en gris.
2. Probabilidad del modelo: barra de progreso y porcentaje numérico que representa la certeza del modelo sobre la clasificación asignada. El color de la barra corresponde a la clasificación (verde, rojo o amarillo).

**Importante:** La clasificación y la probabilidad son generadas exclusivamente por el modelo de inteligencia artificial y no pueden ser modificadas manualmente. Estos valores son de carácter orientativo y deben interpretarse siempre en conjunto con el criterio clínico del especialista.

### ***B.6.3 Historial fotográfico del paciente***

Cada imagen cargada queda almacenada en el historial fotográfico del caso. Las imágenes previas se presentan como miniaturas en la parte inferior del visor principal. Si una imagen tiene una probabilidad asignada por el modelo, esta se muestra superpuesta sobre la miniatura.

Para visualizar una imagen anterior en el visor principal, haga clic sobre su miniatura correspondiente. La barra de probabilidad se actualizará automáticamente para reflejar el resultado específico de esa imagen.

### ***B.6.4 Edición de los datos diagnósticos***

31. Con un caso seleccionado, haga clic en el botón Editar (ícono de lápiz) en el encabezado del panel de diagnóstico.
32. En modo edición podrá modificar: la descripción de la lesión, la fecha del diagnóstico y el estado de confirmación del diagnóstico.
33. Haga clic en Guardar para confirmar los cambios o en Cancelar para descartarlos.

### ***B.6.5 Confirmación del diagnóstico***

Una vez revisados los resultados del modelo y evaluada clínicamente la lesión, el especialista puede registrar su validación diagnóstica activando la casilla de verificación Diagnóstico confirmado por el doctor, disponible en el modo edición del panel de diagnóstico. Esta confirmación queda registrada en el historial del caso y se visualiza en verde en modo lectura.

**Nota:** La confirmación del diagnóstico por parte del especialista es un paso opcional pero recomendado, ya que permite distinguir entre los resultados preliminares del modelo y el diagnóstico clínico validado.

## **B.7 Módulo de perfil de usuario**

El módulo de perfil permite al especialista consultar y actualizar su información de cuenta. Para acceder, haga clic sobre su nombre o avatar en la barra de navegación superior y seleccione Ver perfil o Editar perfil en el menú desplegable.

### ***B.7.1 Información disponible***

3. ID de usuario: identificador único asignado automáticamente por el sistema. Solo lectura.
4. Nombre completo: editable por el especialista.
5. Correo electrónico: editable por el especialista.
6. Fecha de registro: fecha en que se creó la cuenta. Solo lectura.

### ***B.7.2 Actualización del perfil***

34. En la página de perfil, haga clic en el botón Editar.
35. Modifique el nombre o el correo electrónico según sea necesario.
36. Haga clic en Guardar. El sistema actualizará la información y mostrará una confirmación en verde.

## **APÉNDICE C. RESULTADOS DE LAS PRUEBAS DE USABILIDAD**

### **C.1 Introducción**

El presente apéndice documenta los resultados obtenidos durante la sesión de pruebas de usabilidad del prototipo web DermaScan, llevada a cabo con usuarios potenciales del sistema. El objetivo de esta evaluación fue medir la facilidad de uso, la claridad de la interfaz y la percepción clínica del sistema mediante la aplicación del Formulario de Evaluación de Usabilidad descrito en el Apéndice A.

La evaluación se estructuró en cuatro componentes: (1) datos del perfil del participante, (2) escala de usabilidad del sistema SUS, (3) evaluación por tareas guiadas y (4) percepción clínica con preguntas abiertas y afirmaciones de valoración. Los resultados se presentan de forma individual dado el número de participantes en esta primera ronda de pruebas, y se analizan en relación con los criterios de aceptación definidos en los requerimientos no funcionales del sistema.

### **C.2 Perfil del participante**

La prueba de usabilidad fue realizada con un participante cuyo perfil corresponde al actor principal del sistema. A continuación se presenta un resumen de los datos demográficos y de experiencia digital recopilados en la sección 1 del formulario.

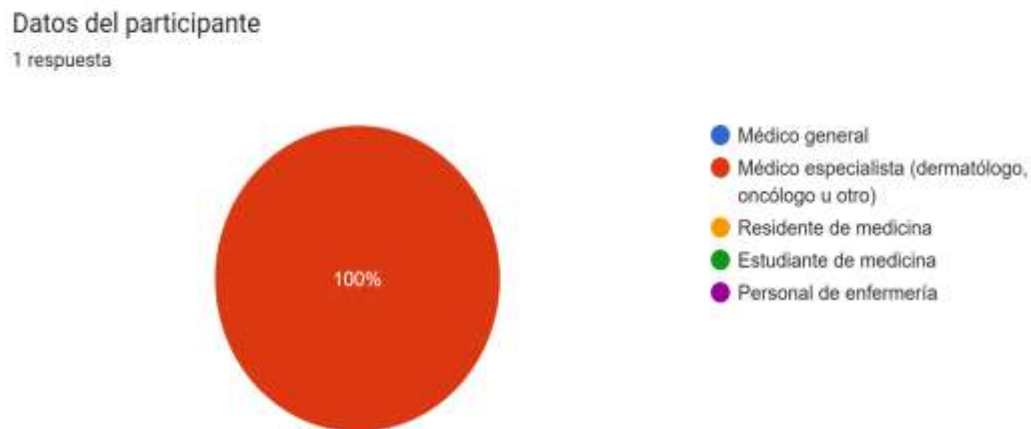
Variable	Respuesta
Ocupacion / rol	Medico especialista (dermatologo, oncologo u otro)
Frecuencia de uso de herramientas digitales en practica medica	Frecuentemente (casi todos los dias)
Experiencia previa con sistemas de apoyo diagnostico basados en IA	No

*Nota. Datos recopilados en la Sección 1 del Formulario de Evaluación de Usabilidad — DermaScan (Apendice A).*

El participante es un profesional de la salud con experiencia en el uso cotidiano de herramientas digitales pero sin exposición previa a sistemas de inteligencia artificial aplicados al diagnóstico médico, lo que representa un escenario de usuario real relevante para validar la curva de aprendizaje del sistema.

### Figura 23.

*Respuestas de la Sección 1 del Formulario de Evaluación de Usabilidad de DermaScan*



## C.3 Resultados de la escala de usabilidad del sistema (SUS)

La Sección 2 del formulario aplicó la escala de usabilidad del sistema SUS (Brooke, 1996), compuesta por diez afirmaciones con valoración del 1 al 5. El cálculo del puntaje SUS sigue el procedimiento estandarizado: para los ítems impares (1, 3, 5, 7, 9) se resta 1 al valor asignado; para los ítems pares (2, 4, 6, 8, 10) se calcula 5 menos el valor asignado. La suma de las contribuciones se multiplica por 2.5 para obtener un puntaje final en escala de 0 a 100.

*Nota.* *Calculo SUS siguiendo el procedimiento de Brooke (1996). Ítems impares: contribucion = valor asignado - 1. Ítems pares: contribución = 5 - valor asignado. Suma total = 36; puntaje final =  $36 \times 2.5 = 90.0$ .*

El puntaje SUS obtenido fue de 90.0 sobre 100, lo que corresponde a la categoría Excelente (grado A) según los rangos de interpretación establecidos en la literatura. Según Bangor, Kortum y Miller (2008), una puntuación superior a 85 indica un nivel de usabilidad considerado como el mejor posible, superando ampliamente el umbral de aceptabilidad (68 puntos) y el umbral de buena usabilidad (71 puntos).

Rango SUS	Calificación	Grado	Resultado Sistema Propuesto
> 85	Excelente	A	90.0 ←
71 - 85	Buena	B / C	
52 - 70	Aceptable	D	
< 52	Deficiente	F	

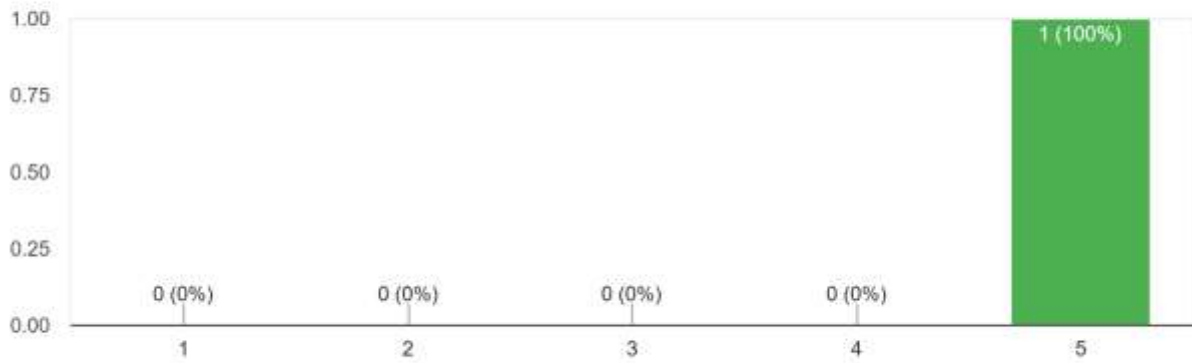
*Nota.* *Rangos de interpretación del puntaje SUS según Bangor, Kortum y Miller (2008).*

**Figura 24.**

*Respuestas de la Sección 2 del Formulario de Evaluación de Usabilidad de DermaScan (Escala SUS).*

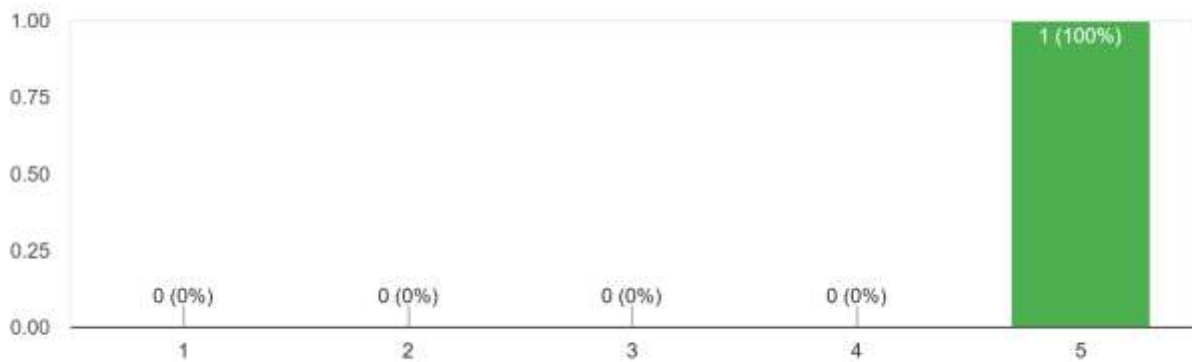
Creo que me gustaría usar este sistema con frecuencia.

1 respuesta



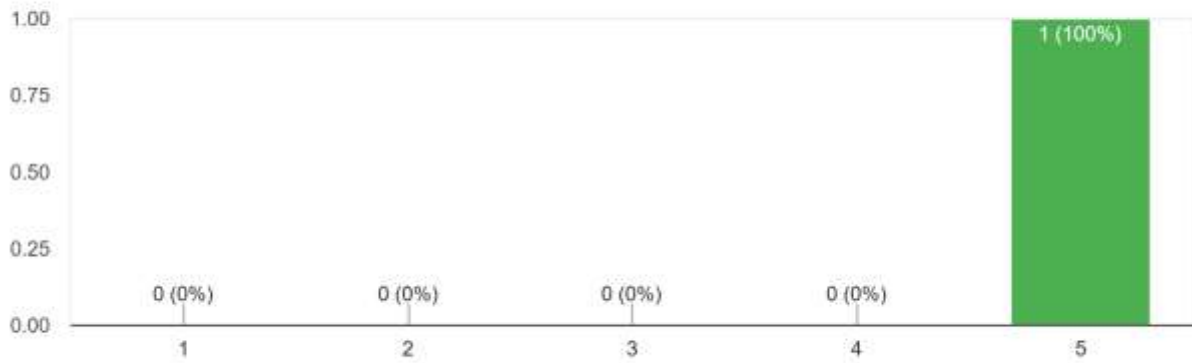
Encontré el sistema innecesariamente complejo.

1 respuesta



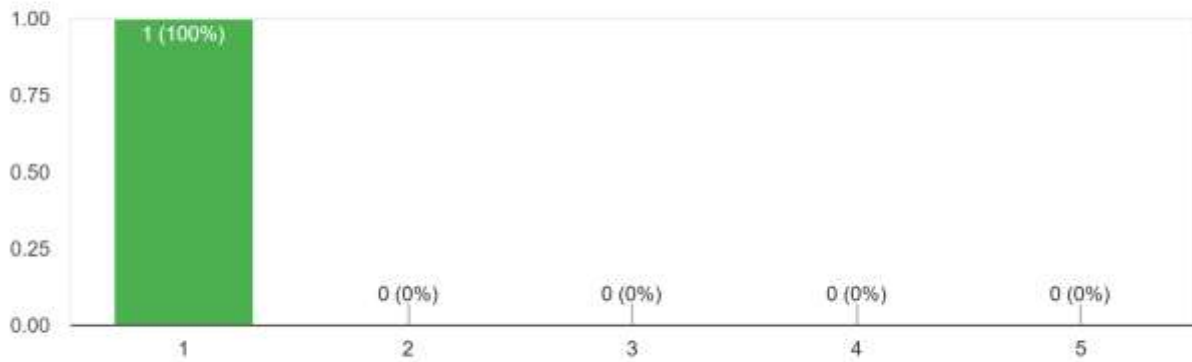
Pensé que el sistema era fácil de usar.

1 respuesta



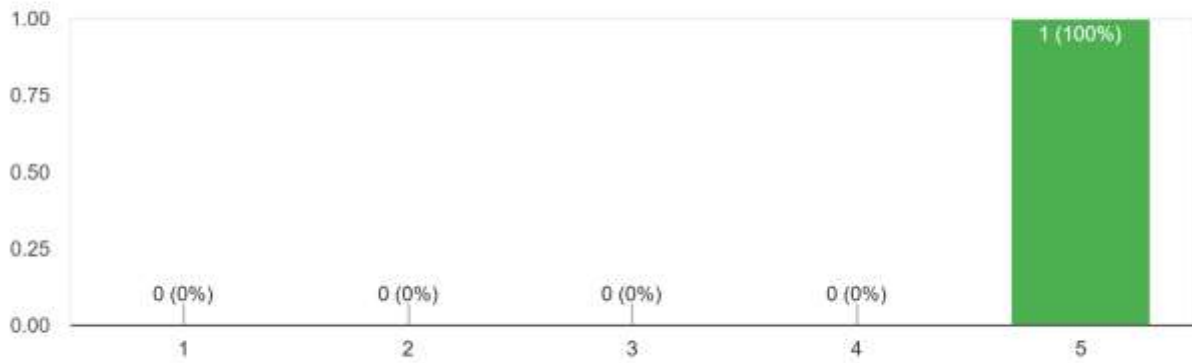
Creo que necesitaría el apoyo de un técnico para poder usar este sistema.

1 respuesta



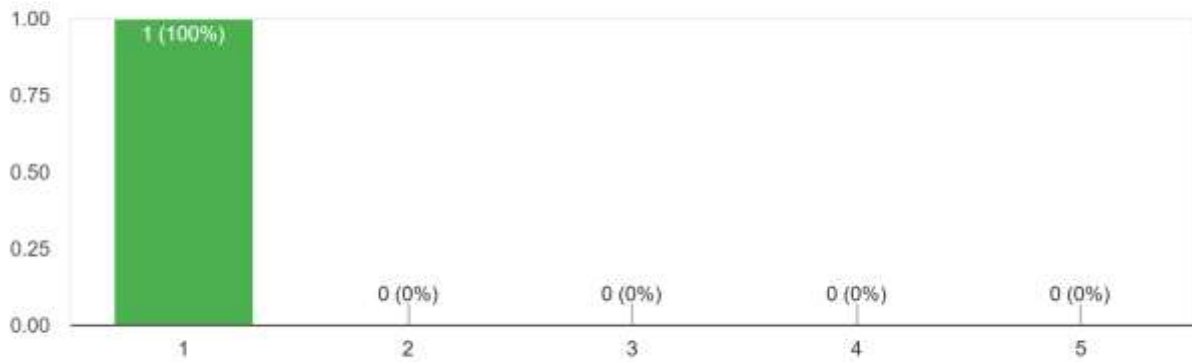
Encontré que las distintas funciones del sistema estaban bien integradas.

1 respuesta



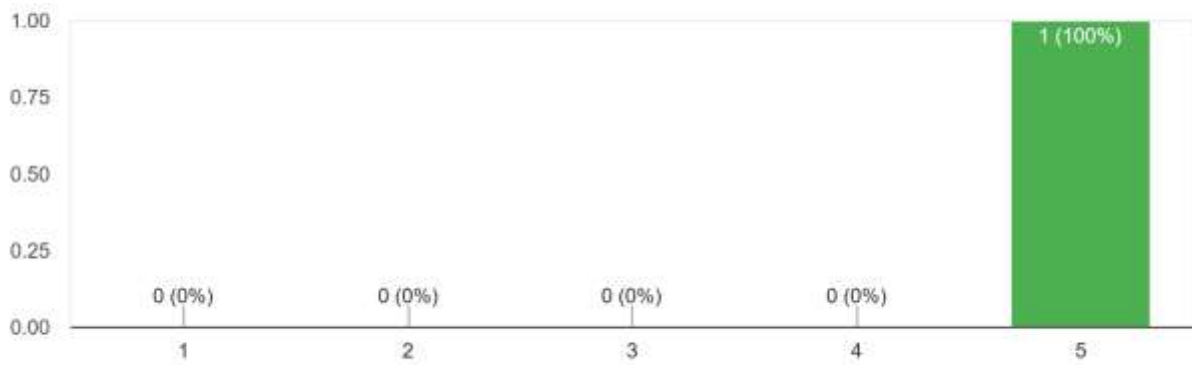
Pensé que había demasiada inconsistencia en el sistema.

1 respuesta



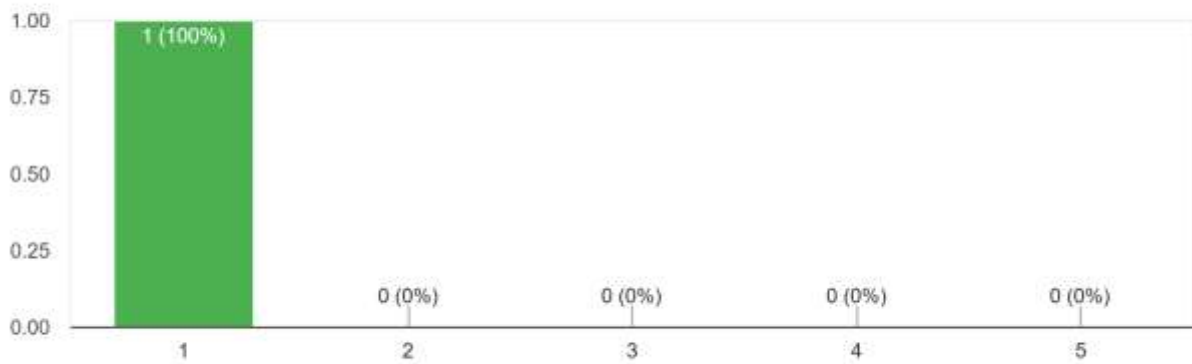
Imagino que la mayoría de las personas aprenderían a usar este sistema muy rápidamente.

1 respuesta



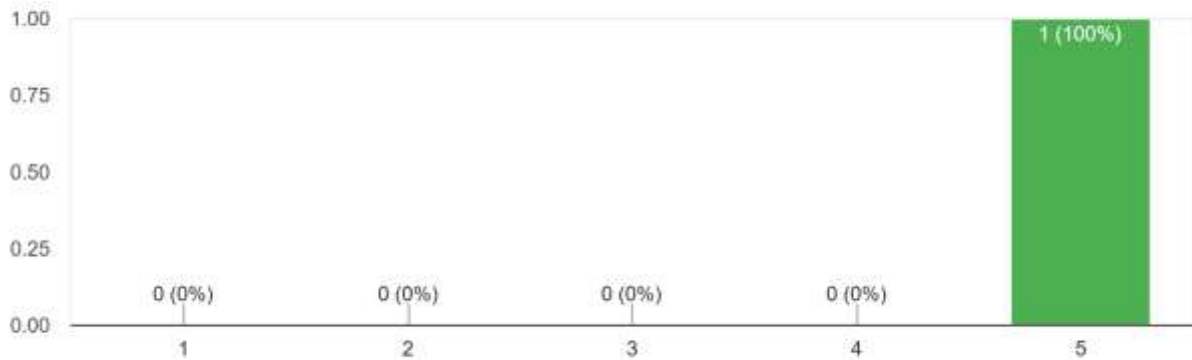
Encontré el sistema muy difícil de manejar.

1 respuesta



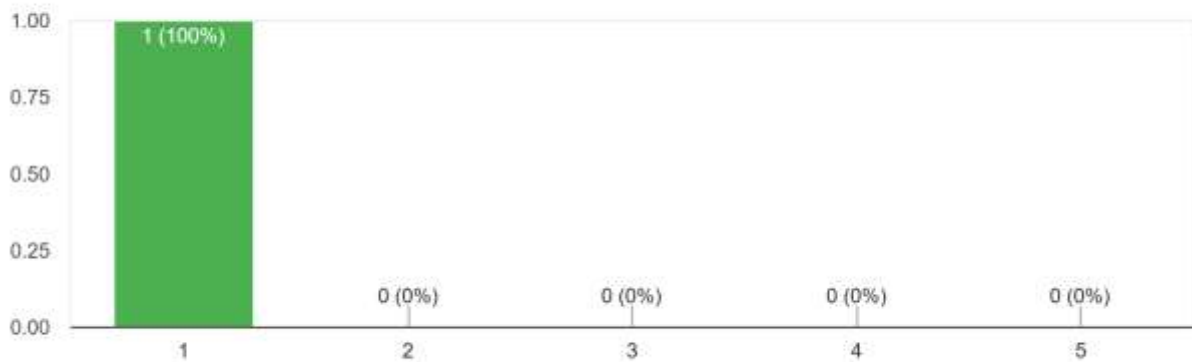
Me sentí muy seguro usando el sistema.

1 respuesta



Necesité aprender muchas cosas antes de poder empezar a usar el sistema.

1 respuesta



#### C.4 Evaluación por tareas

La Sección 3 del formulario evalúa el desempeño del participante en seis tareas guiadas que cubren los principales casos de uso del sistema. Para cada tarea se registró si pudo completarse sin dificultad, el tiempo aproximado de realización y los comentarios espontáneos del participante.

Tarea	Nombre	Completada	Tiempo	Comentario	
1	Registro e inicio de sesión	Si, sin dificultad	ninguna	Entre 30 segundos y 1 minuto	Me parece excelente y fácil de usar.
2	Creación de caso clínico	Si, sin dificultad	ninguna	Entre 30 segundos y 1 minuto	Muy didactico y practico.
3	Carga de imagen dermatoscopia	Si, sin dificultad	ninguna	Entre 1 y 3 minutos	Se pueden cargar sin dificultad.
4	Visualización del resultado diagnóstico	Si, sin dificultad	ninguna	Entre 1 y 3 minutos	Rápido y sin problemas.
5	Confirmación del diagnóstico	Si, sin dificultad	ninguna	Entre 1 y 3 minutos	Acertado.
6	Gestión del perfil de usuario	Si, sin dificultad	ninguna	Entre 1 y 3 minutos	Fácil de realizar.

*Nota. Datos recopilados en la Sección 3 del Formulario de Evaluación de Usabilidad de DermaScan.*

El participante completó las seis tareas propuestas sin dificultad alguna. Cuatro de las seis tareas fueron realizadas en un tiempo de entre uno y tres minutos, mientras que las dos primeras — registro e inicio de sesión y creación de caso clínico — se completaron en menos de un minuto, lo que evidencia una curva de aprendizaje reducida para los flujos más frecuentes del sistema. Ninguna tarea requirió asistencia del equipo de desarrollo ni presentó bloqueos que impidieron su realización.

### **C.5 Percepción clínica y utilidad del sistema**

La Sección 4 del formulario evaluó cinco afirmaciones específicas sobre la utilidad clínica de DermaScan mediante una escala de 1 (totalmente en desacuerdo) a 5 (totalmente de acuerdo).

Estas afirmaciones se diseñaron para capturar dimensiones que van más allá de la usabilidad general y que están directamente relacionadas con la aplicabilidad del sistema en contextos de diagnóstico dermatológico.

Afirmación	Valoración
La visualización de la clasificación (Benigno/Maligno) y el porcentaje de probabilidad es clara y comprensible.	5 / 5
El historial fotográfico del paciente me permite hacer un seguimiento adecuado de la evolución de la lesión.	5 / 5
El mecanismo para confirmar el diagnóstico sugerido por la IA es intuitivo y práctico.	5 / 5
El sistema transmite confianza suficiente para ser considerado como herramienta de apoyo en la práctica clínica.	5 / 5
La distribución de los paneles (historial, ficha del paciente y diagnóstico) facilita el flujo de trabajo clínico.	5 / 5

*Nota. Escala de valoración: 1 = Totalmente en desacuerdo, 5 = Totalmente de acuerdo. Datos recopilados en la Sección 4 del formulario.*

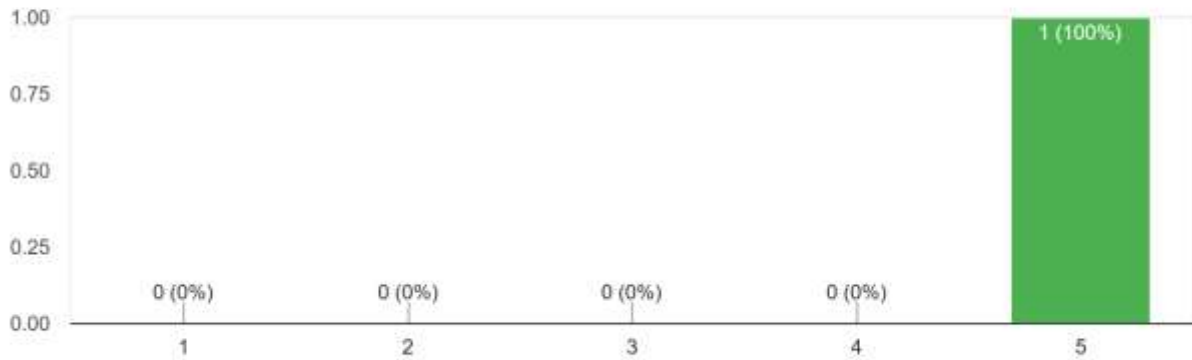
El participante asignó la máxima puntuación (5/5) a las cinco afirmaciones de percepción clínica. Esta evaluación indica que el especialista considera que la visualización de los resultados del modelo es clara y comprensible, que el historial fotográfico facilita el seguimiento de la evolución de las lesiones, que el mecanismo de confirmación diagnóstica es intuitivo, que el sistema transmite suficiente confianza para uso clínico de apoyo, y que la distribución de los paneles de la interfaz es coherente con el flujo de trabajo del especialista.

**Figura 25.**

*Respuestas de la Sección 4 del Formulario de Evaluación de Usabilidad de DermaScan*

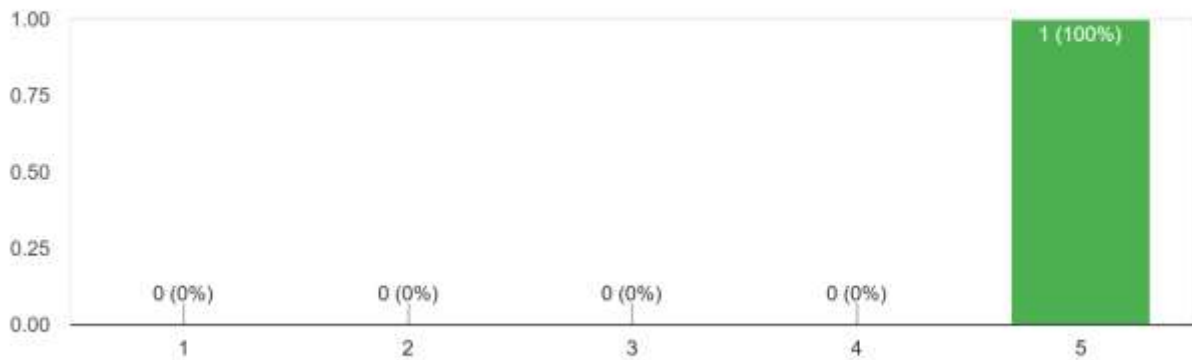
La visualización de la clasificación (Benigno/Maligno) y el porcentaje de probabilidad es clara y comprensible.

1 respuesta



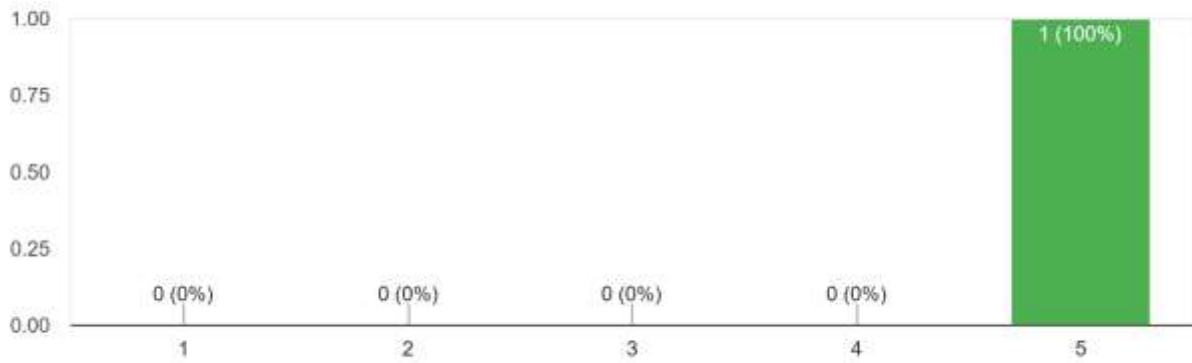
El historial fotográfico del paciente me permite hacer un seguimiento adecuado de la evolución de la lesión.

1 respuesta



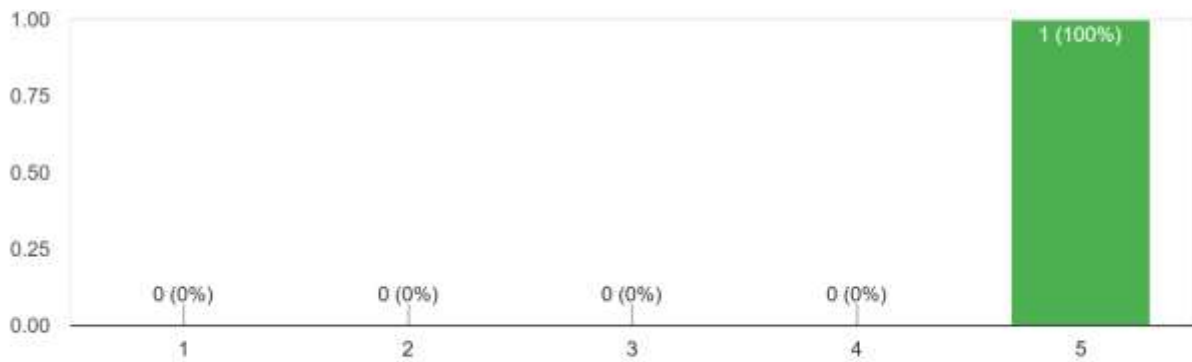
El mecanismo para confirmar el diagnóstico sugerido por la IA es intuitivo y práctico

1 respuesta



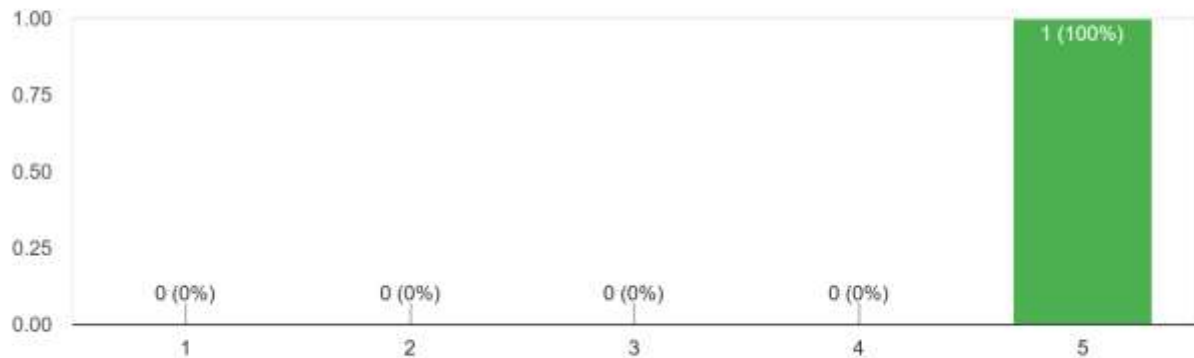
El sistema transmite confianza suficiente para ser considerado como herramienta de apoyo en la práctica clínica.

1 respuesta



La distribución de los paneles (historial, ficha del paciente y diagnóstico) facilita el flujo de trabajo clínico.

1 respuesta



## C.6 Respuestas a preguntas abiertas

La Sección 5 del formulario incluyó cinco preguntas abiertas orientadas a capturar la perspectiva cualitativa del participante sobre la utilidad, las dificultades y las oportunidades de mejora del sistema. A continuación se presentan las respuestas obtenidas acompañadas de un análisis breve.

### C.6.1 ¿Qué aspecto del sistema le resultó más útil o intuitivo?

"Incluir las imágenes porque revisa en tiempo real."

El participante destacó la funcionalidad de carga de imágenes como el aspecto más útil, señalando que permite revisar la lesión en tiempo real durante la consulta. Esta percepción valida el diseño del panel de diagnóstico como el núcleo funcional del sistema y confirma que la integración de imagen y resultado diagnóstico en una misma vista responde a una necesidad clínica real.

### C.6.2 ¿Qué aspecto del sistema le generó más confusión o dificultad?

"Ninguno."

La ausencia de dificultades reportadas en el uso del sistema es consistente con los resultados de la escala SUS (90.0) y con la tasa de completitud del 100% en todas las tareas guiadas. Este hallazgo sugiere que el diseño de la interfaz logró sus objetivos de usabilidad para el perfil del usuario evaluado.

C.6.3 ¿Considera que la presentación de los resultados del modelo es suficientemente clara para orientar una decisión clínica?

"Son suficientes."

El participante consideró que la presentación de los resultados (clasificación y probabilidad) es suficiente para orientar una decisión clínica. Esta valoración es relevante porque valida el enfoque de mostrar simultáneamente la clasificación binaria (benigno/maligno) con su badge de color y la barra de probabilidad numérica, sin sobrecargar la interfaz con información adicional.

C.6.4 ¿Utilizaría este sistema como herramienta de apoyo en su práctica clínica habitual?

"Sí, se puede usar; ayuda a aclarar dudas que se pueden tener en el momento de la valoración."

La respuesta confirma la disposición del especialista a incorporar DermaScan como herramienta complementaria en su práctica clínica, destacando su utilidad para resolver dudas en el momento de la valoración. Esta percepción es coherente con el objetivo del sistema de actuar como apoyo al criterio clínico y no como reemplazo del especialista.

C.6.5 ¿Tiene alguna sugerencia de mejora para el sistema?

"Excelente, seguir usándolo."

La sugerencia de continuar usando el sistema tal como está diseñado, sin proponer cambios concretos, indica un nivel de satisfacción alto con el prototipo en su versión actual. No obstante, se recomienda ampliar el número de participantes en rondas futuras para identificar oportunidades de mejora que no sean evidentes con un único evaluador.

### C.7 Análisis general de los resultados

Los resultados de las pruebas de usabilidad del prototipo DermaScan presentan un panorama consistente y favorable en todas las dimensiones evaluadas. El puntaje SUS de 90.0 ubica al sistema en la categoría Excelente, lo que significa que supera ampliamente los umbrales de usabilidad aceptable (68) y buena (71) reportados en la literatura de referencia (Bangor et al., 2008).

La completitud del 100% en las seis tareas guiadas, sin requerir asistencia técnica en ninguna de ellas, confirma que el sistema es autoexplicativo para un usuario con perfil de especialista médico con experiencia en herramientas digitales. Los tiempos de realización registrados — inferiores a un minuto para los flujos más frecuentes y entre uno y tres minutos para las tareas de mayor complejidad — son coherentes con el requerimiento no funcional RNF02, que establecía que el flujo de diagnóstico debía poder completarse en menos de tres interacciones.

La valoración máxima (5/5) en las cinco afirmaciones de percepción clínica refuerza que el diseño del panel de diagnóstico, en particular la presentación de la clasificación binaria con su badge de color y la barra de probabilidad numérica, es adecuado para el contexto de uso previsto. La disposición de los tres paneles principales (historial de casos, ficha del paciente y panel de diagnóstico) fue valorada positivamente como facilitadora del flujo de trabajo clínico, validando la decisión de arquitectura de información adoptada en el diseño del frontend.

Las respuestas a las preguntas abiertas no identificaron áreas de confusión ni sugerencias de cambio en la interfaz, lo que es consistente con el perfil del participante y con los resultados cuantitativos. No obstante, se reconoce como limitación de esta primera ronda de evaluación el número reducido de participantes ( $n = 1$ ), por lo que los resultados deben interpretarse como indicativos y no como definitivos. Se recomienda realizar una segunda ronda de pruebas con una muestra más amplia — que incluya médicos generales, residentes y personal de enfermería — para identificar posibles brechas de usabilidad que no sean evidentes en el perfil de experto digital evaluado.

<b>Dimensión evaluada</b>	<b>Resultado</b>
Puntaje SUS	<b>90.0 / 100 — Excelente (A)</b>
Tasa de completitud de tareas	<b>6 / 6 tareas completadas (100%)</b>
Tareas sin asistencia técnica	<b>6 / 6 (100%)</b>
Percepción clínica promedio	<b>5.0 / 5.0 en las cinco afirmaciones</b>
Dificultades o bloqueos reportados	<b>Ninguno</b>
Disposición a usar el sistema clínicamente	<b>Si, como herramienta de apoyo</b>

*Nota.* Resumen de los resultados de las pruebas de usabilidad del prototipo DermaScan v1.0.

**Referencias del apéndice**

Bangor, A., Kortum, P. T., y Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574-594.

<https://doi.org/10.1080/10447310802205776>

Brooke, J. (1996). SUS: A quick and dirty usability scale. En P. W. Jordan, B. Thomas, B. A. Weerdmeester, y A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189-194). Taylor & Francis.