

Beyond Transfer Learning: A Lightweight Convolutional Architecture for Dermatoscopic Skin Cancer Detection

Sebastian Lopez¹; Nicolas Navarro²

^{1,2}Universidad Finis Terrae, Chile, slopez.scd@gmail.com, nnavarro2@uft.edu

Abstract– Skin cancer, specifically melanoma, represents a growing public health challenge in Chile, with a 116% increase in the mortality rate over the last two decades. Early detection is critical, but clinical diagnostic accuracy varies significantly, and access to specialists is limited. This work presents the development of an artificial intelligence model for clinical decision support based on deep learning for the automatic classification of skin lesions (benign vs. malignant). The performance of a custom-built Convolutional Neural Network (CNN) architecture, trained from scratch, was compared to a transfer learning model based on InceptionV1. The experimental results indicated that, while the transfer learning model achieved greater overall sensitivity, the proposed architecture attained superior accuracy (69.75% vs. 67.33%), demonstrating a greater capacity to reduce false positives. This validates the effectiveness of designing lightweight and specialized architectures, which achieve competitive and efficient performance without relying on massive pre-training, opening new avenues for the implementation of computer-assisted diagnostic tools in environments with limited computational resources.

Keywords-- Deep Learning, Convolutional Neural Networks, Melanoma, Dermatoscopy.

Más Allá Del Aprendizaje Por Transferencia: Una Arquitectura Convolutiva Ligera Para La Detección Dermatoscópica Del Cáncer De Piel

Sebastian Lopez¹; Nicolas Navarro²

^{1,2}Universidad Finis Terrae, Chile, slopez.scd@gmail.com, nnavarro2@uft.edu

Resumen– El cáncer de piel, específicamente el melanoma, representa un desafío creciente para la salud pública en Chile, con un aumento del 116% en la tasa de mortalidad en las últimas dos décadas. La detección temprana es crítica, pero la precisión diagnóstica clínica varía significativamente y el acceso a especialistas es limitado. Este trabajo presenta el desarrollo de un modelo de inteligencia artificial de soporte a la decisión clínica basado en aprendizaje profundo (Deep Learning) para la clasificación automática de lesiones cutáneas (benignas vs. malignas). Se comparó el rendimiento de una arquitectura de Redes Neuronales Convolucionales (CNN) personalizada y entrenada desde cero, frente a un modelo de aprendizaje por transferencia (Transfer Learning) basado en InceptionV1. Los resultados experimentales indicaron que, si bien el modelo de aprendizaje por transferencia obtuvo una mayor sensibilidad global, la arquitectura propuesta alcanzó una precisión superior (69.75% frente a 67.33%), demostrando una mayor capacidad para reducir los falsos positivos. Esto valida la eficacia del diseño de arquitecturas ligeras y especializadas, las cuales logran un rendimiento competitivo y eficiente sin depender de pre-entrenamientos masivos, abriendo nuevas vías para la implementación de herramientas de diagnóstico asistido en entornos con recursos computacionales limitados.

Palabras clave– Aprendizaje Profundo, Redes Neuronales Convolucionales, Melanoma, Dermatoscopia.

I. INTRODUCCIÓN

El diagnóstico temprano y preciso de lesiones cutáneas malignas es un desafío crítico en la biomedicina actual. El melanoma, aunque representa una fracción menor de los cánceres de piel, es responsable de la gran mayoría de las muertes relacionadas con esta patología. En Chile, la situación es alarmante: según datos del Departamento de Estadísticas e Información de Salud (DEIS), la tasa de defunciones por cáncer de piel aumentó un 116,85% entre 2002 y 2022 [1].

El problema central radica en la variabilidad de la precisión diagnóstica, la cual oscila entre el 24% y el 70% [2] cuando es realizada por personal no especialista en dermatología, sumado a la escasez de dermatólogos (2.9 por cada 100.000 habitantes en Chile), concentrados mayoritariamente en zonas metropolitanas.

Esta investigación propone como objetivo el validar una arquitectura propia, diseñada bajo principios de eficiencia computacional, evitando el uso de aprendizaje por transferencia en la fase de entrenamiento. Se busca contrastar el desempeño técnico del diseño frente a GoogleNet (Inception V1) [3] para determinar si la especialización exclusiva en patrones dermatoscópicos es suficiente para competir con el conocimiento del aprendizaje por transferencia, determinando si un modelo ligero logra un equilibrio aceptable entre sensibilidad y especificidad que minimice las falsas alarmas.

A. Limitaciones

Una limitante central del estudio reside en la utilización exclusiva de bases de datos de acceso gratuito, como el ISIC Archive [4]. Estos repositorios, si bien son fundamentales para la investigación, suelen presentar sesgos demográficos inherentes, caracterizados por una predominancia de fototipos de piel claros (caucásicos) y grupos etarios específicos. Esta falta de representatividad de la diversidad poblacional real puede comprometer la capacidad de generalización del modelo, afectando potencialmente la equidad y precisión del diagnóstico cuando se aplica a pacientes con tonos de piel más oscuros o características dermatológicas locales no representadas en el conjunto de entrenamiento.

Adicionalmente, existe una brecha técnica entre las condiciones idealizadas del entrenamiento y la realidad operativa de la atención primaria. El modelo utiliza imágenes de alta resolución y condiciones controladas, lo cual difiere de la variabilidad de captura encontrada en entornos clínicos reales con recursos limitados. Asimismo, el alcance del sistema se restringe a una clasificación binaria (maligno vs. benigno); esta aproximación no permite la identificación específica de patologías (como diferenciar un carcinoma de un melanoma), limitando la información disponible para la priorización terapéutica por parte del especialista.

II. ESTADO DEL ARTE

En este estado del arte se referencia el conocimiento, técnicas y tecnologías más recientes y avanzadas que existen en la clasificación de imágenes dermatoscópicas mediante Redes Neuronales Convolucionales (CNN).

A. Soluciones Previas en el Mercado Actual

En el mercado actual existen diversas soluciones tecnológicas para el diagnóstico asistido, aunque ninguna ha logrado consolidar una eficacia universalmente alta. En el ámbito del hardware, la "Dermatoscopia digital" se aplica mediante equipos físicos como FotoFinder y Molexmax, capaces de amplificar, almacenar y analizar imágenes para seguimiento temporal. A nivel de software, programas especializados utilizan algoritmos clásicos basados en reglas clínicas como "CHAOS & CLUES" o la regla ABCD, evaluando características como asimetría y bordes como apoyo al especialista. Centrándose en el paciente, aplicaciones móviles como FotoSkin permiten la monitorización personal. Aunque fomentan la detección temprana y el autoexamen, estas herramientas suelen utilizar algoritmos más sencillos basados en reglas, limitando su capacidad diagnóstica profesional.

B. Tecnologías y Estudios Relevantes

Las tecnologías de análisis de imágenes médicas se dividen clásicamente en segmentación (aislamiento de la lesión mediante umbralización o crecimiento de regiones) y extracción de características (etiquetado manual de propiedades geométricas y cromáticas).

A un nivel avanzado, la clasificación ha transitado desde reglas definidas por especialistas (ABCD) y Machine Learning clásico (SVM, KNN, Random Forest) hacia el aprendizaje profundo [5].

A diferencia de los métodos anteriores que requieren intervención manual, las Redes Neuronales Convolucionales (CNNs) aprenden patrones directamente de los datos. La literatura científica reciente evidencia un progreso significativo en este campo. La Tabla 1 presenta una comparativa sintetizada de los estudios más relevantes revisados, destacando arquitecturas, resultados y hallazgos clave.

Tabla 1. Comparativa sobre estudios relevantes en clasificación dermatoscópica.

Estudio (Año)	Arquitectura Modelo	Rendimiento	Hallazgos y Observaciones
Codella et al. (2018) [6]	Ensembles (ResNet, Inception)	AUC > 0.85	Rendimiento competitivo frente a dermatólogos. Limitado por desbalance de clases en ISIC Archive.
Esteva et al. (2017) [7]	Inception (Transfer Learning)	v3 Comparable expertos	Hito en el campo con 129,000 imágenes. La falta de metadatos clínicos restringe su valor práctico.
Ali (2024) [8]	DenseNet	Acc > 83%	Confirmación de la robustez de arquitecturas densas.
Younas (2025) [9]	Air-CNN (Attention-Based)	Acc 91.6%	Incorporación de mecanismos de atención para mejorar la precisión predictiva.
Remya (2024) [10]	ResNet 152v2 + VGG 16 + MobileNet	+ Acc 99%	Establece el estándar actual. Fusión de tres capas combinando metadatos del paciente con imágenes.

C. Brechas Tecnológicas y Oportunidad de Mejora

Pese a los avances, persisten desafíos críticos. La segmentación automática de lesiones complejas (bordes difusos, artefactos) sigue afectando el desempeño global. Asimismo, la falta de datos longitudinales impide entrenar modelos que reconozcan la evolución temporal, clave en la detección temprana.

Finalmente, existe una oportunidad clara en el desarrollo de frameworks híbridos que combinen Aprendizaje profundo con conocimiento clínico explícito (reglas visuales), aumentando la interpretabilidad y aceptación médica, además de su integración en plataformas escalables para zonas con escaso acceso a especialistas.

III. DESARROLLO

A. Selección del Dataset

Para garantizar la calidad del entrenamiento y evitar inconsistencias en los datos, se realizó la selección del repositorio BCN 20.000 del ISIC Archive [4].

Como se menciona el repositorio utilizado corresponde al BCN 20.000, el cual almacena alrededor de 19.000 imágenes de carácter dermatoscópicas, principalmente fue seleccionado dado su balance de clases Benignas/Malignas, tal como se puede evidenciar en Fig. 1, donde se compara con demás repositorios de imágenes del ISIC Archive [4]:

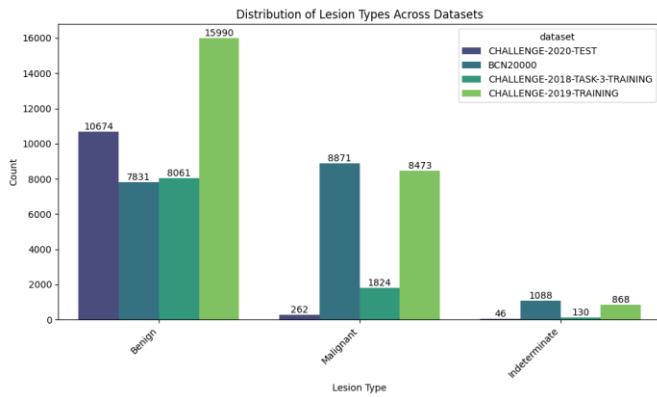


Fig. 1: Distribución por género en 4 dataset.

Luego de realizar la limpieza de metadatos y validación de imágenes, se consolidó un corpus de 11.698 imágenes disponibles para el uso en entrenamiento.

El total de imágenes mencionadas fueron divididas mediante la estratificación del 70% para entrenamiento y 30% para validación interna, velando que la proporción Benigno/Maligno se mantuviera proporcional en ambos subsets. En cuanto al preprocesamiento de las imágenes, dado que presentaban dimensiones y resoluciones heterogéneas, se realizó un reescalado dinámico a un tamaño estándar de 224×224 píxeles, manteniendo el formato RGB, para luego realizar la normalización de estas a un rango $[0, 1]$, mediante el factor de conversión $1./255$. Cabe destacar que estos parámetros de reescalado fueron seleccionados gracias a una revisión del estado del arte, donde los valores estándar de la investigación en la disciplina del aprendizaje profundo son de 224×224 píxeles, dada su compatibilidad con modelos diseñados para realizar aprendizaje por transferencia.

B. Arquitectura del Modelo Propuesto

El modelo propuesto en esta investigación fue diseñado para abordar una clasificación binaria de lesiones cutáneas, enfocándose en la profundidad para la obtención de características y la eficiencia de los parámetros de cada capa, para evitar el sobreajuste. La arquitectura de este modelo se divide en 3 áreas críticas.

En primera instancia se debe a los bloques de extracción de características, se menciona que la red consta de una secuencia de bloques convolucionales diseñados para un aprendizaje jerárquico. Inicialmente mantiene 32 filtros para capturar características de bajo nivel, como los bordes y contrastes de colores. Posterior a este bloque se realiza un aumento de filtros a 64 y 128 filtros en cada bloque respectivo, para la detección de patrones dermatológicos complejos como las estructuras vasculares de las lesiones y redes de pigmentación. Para mantener un aumento progresivo de bloques convolucionales, la configuración de estos es clave para el desempeño necesario

que se mantiene, es por lo que cada etapa convolucional sigue a detalle el siguiente patrón de diseño:

Conv2D (3 x 3) → Batch Normalization → Activación ReLU → Maxpooling2D (2 x 2).

En segunda instancia, la regularización estructural fue clave para mitigar la co-adaptación de las neuronas, el cual es un riesgo alto en este tipo de dataset de origen médico y tamaño mediano, para esto se insertaron capas de *Dropout* con una tasa del 25% (0.25) tras cada bloque extractor de características, forzando al modelo a desarrollar rutas redundantes para la identificación de malignidad, mejorando la robustez del modelo ante imágenes que mantengan artefactos externos o ruido.

Y en última instancia, el uso de un cabezal de clasificación optimizado, a diferencia de las arquitecturas clásicas como la VGG16 [3] es la sustitución de una capa del tipo Plana por una capa que Agrupa el promedio global del vector de características 2D (GAP). Esta decisión se basa en la eficiencia de los parámetros de entrenamiento, debido a que las capas Planas mantienen la ubicación espacial del vector de características masivo, en cambio la capa GAP condensa los mapas de características ($7 \times 7 \times 128$) en un solo vector promedio de 128 valores ($1 \times 1 \times 128$), reduciendo de manera considerable los parámetros de entrenamiento del clasificador a un número cercano de 16.000, disminuyendo el riesgo de sobreajuste.

C. Entrenamiento del Modelo

Debido a que el modelo comienza su entrenamiento desde cero, la solución a esto fue el diseño e implementación de un flujo de trabajo de entrenamiento específico para maximizar el poder de generalización del modelo. Lo primero que se realizó fue un Aumento de datos agresivo, donde se aplicaron transformaciones geométricas en tiempo real para simular la variabilidad clínica, a continuación, se consolida en la Tabla 2:

Tabla 2. Rangos de aumento de datos aplicados.

Transformación	Rango Aplicado
Rotación	-30%, + 30%
Traslación	20%
Cizallamiento	20°
Zoom	0.8 – 1.2
Volteos Verticales	Sí
Volteos Horizontales	Sí

En cuanto al optimizador utilizado, este corresponde al RMSprop ($lr = 0.0005$, $P=0.92$), debido a su capacidad de adaptar la tasa de aprendizaje en función de la media móvil de los gradientes recientes, junto a este optimizador se implementó un algoritmo de reducción de la tasa de aprendizaje, implementando el ReduceLROnPlateau, disminuyendo la tasa de aprendizaje si la precisión de la validación no mejora pasadas 5 épocas.

El hardware utilizado para esta investigación fue una GPU Nvidia Tesla L4 de 24 Gb de memoria VRAM, con una duración de 100 épocas de entrenamiento, este valor fue seleccionado a través de cinco experimentos realizados, donde la cantidad de 100 épocas fue la que mejores resultados obtuvo.

D. Resultados de Entrenamiento

Luego de realizar el entrenamiento con todos los parámetros mencionados, se realizó la gráfica del rendimiento de Exactitud y Pérdida en el entrenamiento, la cual entrega obteniendo la gráfica que se observa en la Fig. 2:



Fig. 2. Resultados del Entrenamiento Modelo Propuesto.

La Fig. 2 evidencia la convergencia del modelo, observándose una disminución progresiva de la pérdida y un aumento sostenido de la exactitud hasta alcanzar la estabilidad, lo que confirma que las estrategias de regularización y aumento agresivo de datos disminuyeron las posibilidades de sobreajuste, demostrando que la red logra diferenciar de manera aceptable los patrones dermatológicos de las imágenes de entrenamiento, en lugar de memorizar el set de entrenamiento. Mientras que, en la validación interna del entrenamiento, al graficar el Val-Loss y Val-Accuracy, se obtiene la Fig. 3:

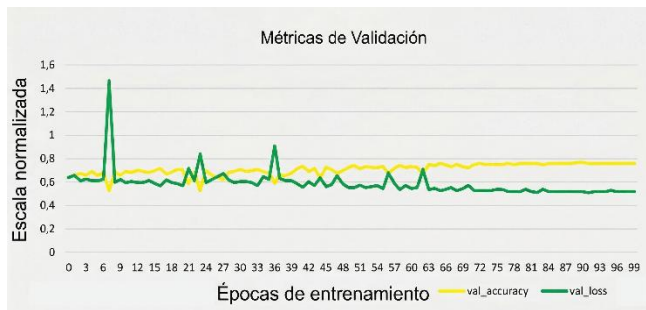


Fig. 3. Resultados de Validación Interna Modelo Propuesto.

En cambio, la Fig. 3 evidencia el rendimiento de la exactitud y pérdida del modelo en un set de datos no utilizado para el entrenamiento, siendo una prueba de que es funcional este sistema para lesiones que están fuera de los datos de entrenamiento.

Para realizar el cálculo de las métricas estándar de rendimiento del modelo, se utilizó un dataset externo al del entrenamiento, con el fin de que sean apreciaciones reales y objetivas, para esto se usó el ISIC Challenge 2019 [4], el cual consta de 24.463 imágenes, donde 15.990 imágenes corresponden a lesiones Benignas (Nevus, Dermatofibroma, Lentigo solar, Cicatriz y Queratosis Seborreica) y 8.473 a lesiones Malignas (Melanoma (NOS), Carcinoma Basocelular, Carcinoma de Células Escamosas (NOS) y Melanoma Metastásico) con un desbalance casi del 50% a favor de las lesiones benignas.

Las métricas calculadas corresponden a la Precisión, Especificidad, F1-Score y Sensibilidad, tal como se muestra en la Tabla 3:

Tabla 3. Métricas calculadas del Modelo Propuesto.

Métrica Evaluación	de Valor Obtenido	Descripción del Desempeño
Precisión (Precision)	69,75%	Alta fiabilidad en la predicción de casos positivos, minimizando falsas alarmas.
Especificidad	80,87%	Excelente capacidad para descartar correctamente lesiones benignas (sanos).
F1-Score	60,49%	Balance armónico robusto entre la precisión y la recuperación de casos.
Sensibilidad (Recall)	53,40%	Capacidad moderada de detección, priorizando la certeza sobre la cobertura total.

El cálculo de las métricas menciona que el modelo propuesto mantiene una precisión cercana al 70% y una especificidad del 80%. Esto indica que el modelo logra descartar de manera aceptable las lesiones benignas, es decir, separa de manera consistente las lesiones malignas de las benignas.

E. Análisis Comparativo

Para validar el modelo propuesto, se replicaron las condiciones experimentales y arquitectura de modelo, utilizada en la literatura, específicamente en el estudio realizado por Ummature et al. [3], usado como modelo referencial de rendimiento.

En primera instancia, la arquitectura a utilizar corresponde a la GoogleNet (Inception V1), preentrenada con el repositorio ImageNet [11], por lo que se utilizará una estrategia de Aprendizaje por transferencia, congelando la base convolucional completa de dicho modelo (Aproximadamente 5.6 Millones de parámetros), para luego definir un *Dropout* de 0.4 y una capa densa con Salida *Softmax* Binaria (Clasificador Benigno / Maligno).

En cuanto al entrenamiento, se utilizó un optimizador Adam ($\text{lr} = 0.0001$), 25 épocas de entrenamiento y no se realizó el aumento de datos, tal como lo indican en la literatura seleccionada, pero si se realizó la normalización de las imágenes y el reescalado a la resolución de 224×224 píxeles,

manteniendo el formato RGB, para luego realizar la normalización de estas a un rango [0, 1], mediante el factor de conversión 1. /255.

Con el fin de obtener una visión detallada sobre la distribución de aciertos y errores, se generó la matriz de confusión comparativa utilizando el conjunto de datos de prueba y los resultados obtenidos por ambos modelos de inteligencia artificial, tal como muestra la Tabla 4.

Tabla 4. Matriz de Confusión de la Distribución de los Datos de Prueba y Predicciones del Modelo de Transfer Learning junto al Modelo Propuesto.

Matriz de Confusión Dataset Original		
Distribución	Predicción Positiva (Maligna)	Predicción Negativa (Benigna)
Real Positivo (Maligna)	8.473	0
Real Negativo (Benigna)	0	15.990
Matriz de Confusión Transfer Learning		
Distribución	Predicción Positiva (Maligna)	Predicción Negativa (Benigna)
Real Positivo (Maligna)	5.705	3.397
Real Negativo (Benigna)	2.768	12.593
Matriz de Confusión Modelo Propuesto		
Distribución	Predicción Positiva (Maligna)	Predicción Negativa (Benigna)
Real Positivo (Maligna)	5.910	5.157
Real Negativo (Benigna)	2.563	10.833

El modelo propuesto registra un desempeño de 5.910 Verdaderos Positivos junto a 10.833 Verdaderos Negativos, lo que demuestra una identificación efectiva de un gran número de lesiones malignas presentes en la base de datos de control. En contraste a la arquitectura Inception V1 [3] presenta 5.705 Verdaderos Positivos y 12.593 Verdaderos Negativos, el cual muestra una mayor capacidad de descartar casos negativos.

Al analizar la distribución de errores basada en los datos de la matriz de confusión, se observa que el modelo propuesto reduce las falsas alarmas al limitar los Falsos Positivos a 2.563 casos frente a los 2.768 casos identificados por el modelo de aprendizaje por transferencia.

No obstante, dicha configuración presenta una mayor incidencia en la omisión de los casos críticos de 5.157 Falsos Negativos en comparación con los 3.397 casos de la arquitectura de referencia, subrayando el compromiso técnico entre la precisión y sensibilidad inherente a la especialización de cada una de las arquitecturas.

Para que las métricas de rendimiento tuvieran sentido en cuanto a la comparación, se realizó el cálculo de estas utilizando el Repositorio del ISIC Archive, Challenge 2019 [4], con el objetivo de que las diferencias de rendimiento sean adjudicables netamente a la arquitectura y estrategia de entrenamiento. A continuación, se consolida comparativa de rendimiento en la Tabla 5.

Tabla 5. Comparativa de rendimiento Modelo Propuesto frente a Inception V1.

Métrica	Modelo (Propuesto)	2 GoogleNet (Referencia) [3]	Interpretación Técnica
Precisión (Precision)	69,75 %	67,33%	El Modelo 2 discrimina mejor los Falsos Positivos.
Sensibilidad (Recall)	53,40%	62,68 %	GoogleNet recupera más casos positivos (mayor cobertura).
Especificidad	80,87%	81,98 %	Ambos modelos son efectivos descartando sanos.
AUC-ROC	56,95%	63,8 %	Comparativa de rendimiento
F1-Score	60,49%	64,92 %	GoogleNet mantiene un mejor equilibrio global.

Para complementar el análisis de desempeño y permitir una comparativa independiente del umbral de decisión, se incorporó el cálculo del Área bajo la curva (AUC). Esta métrica permite evaluar la capacidad del modelo para discriminar entre clases a lo largo de todo el espectro de sensibilidad y especificidad, proporcionando así una visión más integral de la robustez y capacidad predictiva de ambas arquitecturas.

El hallazgo más relevante de esta comparación corresponde al mejor desempeño del Modelo Propuesto en términos de precisión (+2.42%). Este resultado sugiere que el entrenamiento del modelo desde cero favoreció una identificación más efectiva de texturas específicas de la piel. En contraste, la arquitectura GoogleNet, al utilizar pesos preentrenados obtenidos desde ImageNet [11], mostró una tendencia a presentar una mayor sensibilidad, aunque con una menor especificidad frente a lesiones cutáneas.

Es importante destacar que los valores de rendimiento obtenidos, situados en torno al 70%, difieren de las métricas reportadas en algunos estudios previos. Esto podría explicarse por el rigor metodológico asociado al uso de un conjunto de datos de prueba externos. Esta estrategia contribuye a reducir el sobreajuste y minimizar el sesgo de evaluación que suele producirse cuando los modelos son evaluados utilizando datos pertenecientes al mismo dominio empleado durante el entrenamiento.

IV. CONCLUSIÓN Y DISCUSIÓN

La investigación valida que la arquitectura de redes neuronales convolucionales propuesta, al ser optimizada con una menor cantidad de parámetros de entrenamiento y con un régimen agresivo de Aumento de datos, llega a ser competitiva frente a modelos pre-entrenados.

Si bien, el aprendizaje por transferencia [3] ofrece ventajas de sensibilidad, el modelo propuesto demostró una mejor precisión global, en cuanto al diagnóstico positivo, crucial para reducir los diagnósticos erróneos o la derivación a procedimientos invasivos. Bajo esta premisa, se sugiere que el modelo ligero del diagnóstico clínico, o alternativamente, ser sometido a un ajuste en su umbral de decisión para priorizar la sensibilidad según el escenario clínico requerido.

No obstante, es importante reconocer que, en un contexto oncológico, la mayor sensibilidad demostrada por Inception V1 posee una relevancia clínica. La capacidad de detectar la mayor cantidad posible de lesiones malignas resulta prioritaria para minimizar los falsos negativos, los cuales corresponden a casos de melanoma no diagnosticados. En este sentido, el modelo ligero no debería considerarse superior únicamente en función de su precisión global, sino que su desempeño debe analizarse de manera integral considerando el equilibrio entre sensibilidad, especificidad y el impacto clínico asociado a cada tipo de error diagnóstico.

Estos resultados apoyan la postura de que el diseño de arquitecturas ligeras y especializadas podrían superar a modelos robustos y generalistas en tareas de complejidad significativa, en entornos clínicos específicos. Sin embargo, una diferencia del 2,42% en la precisión es insuficiente para declarar una superioridad algorítmica definitiva. Sin pruebas de significancia estadística, no es posible discernir si la mejora es estructural o producto de la varianza estocástica del entrenamiento, siendo indispensable realizar contrastes de hipótesis para validar la reproducibilidad de los resultados.

Con una vista hacia el futuro, se busca mejorar este modelo ligero, a través del uso de datasets de lesiones dermatoscópicas chilenas y latinoamericanas, mitigando los sesgos demográficos de los repositorios internacionales y asegurando la equidad algorítmica en el diagnóstico local.

Adicionalmente, se proyecta la implementación de técnicas de Ensemble Learning y la integración de metadatos clínicos para maximizar la sensibilidad del sistema, buscando concretar una herramienta robusta que contribuya a mitigar el incremento de la tasa de mortalidad por cáncer de piel en Chile.

REFERENCIAS

- [1] L. J. Cevallos Orellana, "Prevalencia de cáncer de piel en pacientes de 18 a 50 años en el Hospital Teodoro Maldonado Carbo durante el período 2014-2019," Tesis de grado, Univ. de Guayaquil, 2020.
- [2] Y. Liu et al., "A deep learning system for differential diagnosis of skin diseases," *Nature Medicine*, vol. 26, pp. 900–908, 2020.
- [3] S. B. Ummappure, R. Tilekar, and S. Mallappa, "Skin Cancer Classification using VGG-16 and Googlenet CNN Models," *International Journal of Computer Applications*, vol. 184, no. 42, pp. 5–9, 2023.

- [4] ISIC Archive, "International Skin Imaging Collaboration: Melanoma Project," [Online]. Available: <https://www.isic-archive.com>.
- [5] Á. Martínez Martínez, "Análisis de técnicas de aprendizaje profundo aplicadas a la clasificación de imágenes," Trabajo de fin de grado, Univ. Politécnica de Madrid, 2014.
- [6] N. C. F. Codella et al., "Deep learning ensembles for melanoma recognition in dermoscopy images," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1756–1768, 2018.
- [7] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [8] Ali, S. N., Jain, S., Gupta, R., & Acharya, D. (2023). A web-based mpox skin lesion detection system using state-of-the-art deep learning. *Scientific Reports*, 13, Article 12345. <https://doi.org/10.1038/s41598-023-12345-6>
- [9] Younas, S., Sargano, A. B., You, L., & Habib, Z. (2025). Attention-based inception-residual CNN: Skin cancer diagnosis with multiclass classification. *Information*, 16(2), 120. <https://doi.org/10.3390/info16020120>
- [10] Remya, S., Anjali, T., & Sugumaran, V. (2024). A novel transfer learning framework for skin cancer classification using dermoscopic images. *Biomedical Signal Processing and Control*, 83, 104644. <https://doi.org/10.1016/j.bspc.2023.104644>
- [11] ImageNet, "ImageNet Large Scale Visual Recognition Challenge," [Online]. Available: <https://www.image-net.org>.