

Comparación y Clasificación Automática de Lesiones Dermatoscópicas mediante ResNet50 y EfficientNetB3 con Transfer Learning: Implementación sobre el Dataset HAM10000

Comparison and Automatic Classification of Dermoscopic Lesions Using ResNet50 and EfficientNetB3 with Transfer Learning: Implementation on the HAM10000 Dataset

Amin Edoardo Pacheco Álvarez, Gabriel Rivero Gómez

Universidad Modelo — Ingeniería Biomédica

Ingeniería Biomédica, 6.º Semestre — Procesamiento de Imágenes Médicas

RESUMEN

Resumen— El cáncer de piel representa uno de los problemas de salud pública más relevantes a nivel mundial, y su diagnóstico temprano es determinante para mejorar el pronóstico del paciente. En este trabajo se presenta la comparación de dos arquitecturas de clasificación de lesiones dermatoscópicas con Transfer Learning: ResNet50 y EfficientNetB3, ambas preentrenadas con pesos de ImageNet disponibles en Keras Applications. El objetivo fue comparar dos estrategias distintas de Transfer Learning: Fine-Tuning completo en ResNet50, donde se dejaron todas las capas libres para reentrenarse, y Extracción de Características en EfficientNetB3, donde las capas de la red base se mantuvieron congeladas y solo se entrenó el bloque de clasificación final. Ambos modelos fueron evaluados sobre un subconjunto balanceado del dataset HAM10000 con tres tipos de lesiones: nevo melanocítico (nv), melanoma (mel) y queratosis benigna (bkl). Para manejar el desbalance entre clases, se aplicó undersampling a la clase mayoritaria y data augmentation con parámetros justificados en literatura científica a las clases minoritarias, logrando 3,000 imágenes por clase. Los resultados muestran que ResNet50 alcanzó una accuracy de 0.9059 en 30 épocas, mientras que EfficientNetB3 obtuvo 0.8748 en 23 épocas sobre el conjunto de prueba independiente de 1,350 imágenes.

Palabras clave— Clasificación de lesiones dermatoscópicas, ResNet50, EfficientNetB3, Transfer Learning, HAM10000, data augmentation, deep learning, comparación de arquitecturas.

ABSTRACT

Abstract— Skin cancer represents one of the most relevant public health problems worldwide, and its early diagnosis is decisive for improving patient prognosis. This paper presents the comparison of two dermoscopic lesion classification architectures using Transfer Learning: ResNet50 and EfficientNetB3, both pretrained with ImageNet weights and available in official Keras Applications. The main objective was to compare two different Transfer Learning strategies: Full Fine-Tuning on ResNet50, where all layers were unfrozen for retraining, and Feature Extraction on EfficientNetB3, where the base network layers were kept frozen and only the final classification block was trained. Both models were evaluated on a balanced subset of the HAM10000 dataset comprising three diagnostic categories: melanocytic nevi (nv), melanoma (mel) and benign keratosis (bkl). To address dataset imbalance, undersampling was applied to the majority class and data augmentation with clinically justified parameters was applied to minority classes, achieving 3,000 images per class. Results show ResNet50 achieved an accuracy of 0.9059 over 30 epochs, while EfficientNetB3 obtained 0.8748 over 23 epochs on an independent test set of 1,350 images.

Keywords— Dermoscopic lesion classification, ResNet50, EfficientNetB3, Transfer Learning, HAM10000, data augmentation, deep learning, architecture comparison.

I. INTRODUCCIÓN

El melanoma y otras formas de cáncer de piel están entre los tipos de cáncer con mayor incidencia a nivel mundial. Según la OMS, se diagnostican aproximadamente 1.5 millones de nuevos casos de cáncer de piel no melanoma y más de 300,000 casos de melanoma al año [1]. Detectarlo a tiempo hace una diferencia importante: en etapas tempranas la tasa de supervivencia a cinco años supera el 98%, mientras que en etapas avanzadas baja a menos del 25% [2]. La dermoscopia es una técnica de imagen no invasiva que permite visualizar estructuras de la piel que no son visibles a simple vista, mejorando la precisión del diagnóstico. Sin embargo, interpretar estas imágenes requiere experiencia especializada y los resultados pueden variar bastante entre observadores [3]. Debido a esto, en los últimos años se ha incrementado el uso de modelos de deep learning como apoyo al diagnóstico dermatológico. Estos sistemas no buscan reemplazar al especialista, sino ayudar a detectar patrones difíciles de identificar únicamente mediante observación visual.

Las redes neuronales convolucionales (CNN) han mostrado resultados muy buenos en clasificación de imágenes médicas. En particular, arquitecturas como ResNet50 [4] y EfficientNetB3 [5], preentrenadas con ImageNet, han tenido buen desempeño en datasets públicos de la comunidad ISIC, siendo el Transfer Learning permite aprovechar el conocimiento de un modelo entrenado en datos generales y adaptarlo a una tarea más específica, como la clasificación dermatoscópica, sin tener que partir desde cero. Por lo que en este trabajo se comparan ResNet50 y EfficientNetB3 para la clasificación de lesiones dermatoscópicas usando el dataset HAM10000 [6], abordando el desbalance de clases mediante técnicas de balanceo con parámetros de data augmentation justificados en literatura [7, 8].

II. PROBLEMA A RESOLVER

El diagnóstico automático de lesiones de piel mediante dermoscopia enfrenta dos problemas principales: el desbalance entre clases del dataset disponible, y la elección de la arquitectura de deep learning más adecuada para el tamaño y las características del dataset.

El dataset HAM10000 tiene un desbalance bastante notable, siendo que la clase mayoritaria (nevo melanocítico) cuenta con 6,705 imágenes, mientras que melanoma solo tiene 1,113 y queratosis benigna 1,099 y ese desbalance genera problemas durante el entrenamiento, haciendo que el modelo tienda a favorecer la clase con más muestras [9].

Además, en cuanto a la literatura esta reporta resultados contradictorios sobre qué arquitectura es superior para HAM10000: algunos trabajos indican que EfficientNet supera a ResNet50 [10], mientras que otros muestran resultados similares [11]. Por esta razón se decidió comparar ambas arquitecturas bajo las ciertas condiciones experimentales, buscando observar cuál se adapta mejor al dataset HAM10000 y cómo influye la estrategia de entrenamiento utilizada para cada una.

Para abordar esto, se propone: (1) un esquema de balanceo con parámetros de augmentation donde se buscó que sean clínicamente justificados, (2) la implementación de ResNet50 y EfficientNetB3 con Transfer Learning desde Keras Applications, y (3) la comparación de resultados sobre el mismo conjunto de prueba independiente.

III. OBJETIVOS

A. Objetivo General

Comparar el desempeño de las arquitecturas ResNet50 y EfficientNetB3 con Transfer Learning para la clasificación automática de lesiones dermatoscópicas en tres categorías diagnósticas del dataset HAM10000, evaluando accuracy, loss, precisión, recall y F1-score por clase mediante matrices de confusión y reportes de clasificación.

B. Objetivos Específicos

- 1) Preprocesar y balancear el dataset HAM10000 para las clases nv, mel y bkl mediante undersampling y data augmentation con parámetros clínicamente justificados, logrando 3,000 imágenes por clase.
- 2) Implementar ResNet50 con pesos de ImageNet desde Keras Applications oficial utilizando el patrón de Transfer Learning de la guía oficial de TensorFlow.
- 3) Implementar EfficientNetB3 con pesos de ImageNet desde Keras Applications oficial con su preprocesamiento interno.
- 4) Entrenar y evaluar ambos modelos bajo distintas estrategias de Transfer Learning para evaluar el impacto de cada enfoque en el desempeño final de la clasificación.
- 5) Analizar y comparar los resultados con el estado del arte reportado en la literatura para el dataset HAM10000.

IV. ANTECEDENTES

La clasificación de lesiones de piel con imágenes dermatoscópicas es un tema bastante trabajado dentro de la visión computacional médica. A continuación, se revisan los trabajos más relevantes que encontramos para contextualizar este proyecto:

A. ResNet50 aplicado a lesiones de piel

ResNet50 fue propuesta por He et al. [4] en 2015 e introdujo las conexiones residuales que resuelven el problema del gradiente desvaneciente en redes profundas. Con 24,121,219 parámetros totales y pesos preentrenados en ImageNet, ha demostrado ser efectiva para clasificación de lesiones de piel con Transfer Learning. Harangi [12] reportó una accuracy de 86.7% con ResNet50 sobre el dataset ISIC 2017.

B. EfficientNetB3 aplicado a lesiones de piel

En 2019, Tan y Le propusieron EfficientNet [5], una arquitectura enfocada en optimizar el escalado de la red neuronal. A diferencia de otros modelos que solo aumentan profundidad o resolución, EfficientNet ajusta varios parámetros simultáneamente, logrando buenos resultados con menos costo computacional, EfficientNetB3 tiene 11,183,922 parámetros totales, bastante menos que ResNet50, pero alcanza una accuracy de 81.6% en ImageNet frente al 76% de ResNet50, y en estudios recientes sobre HAM10000 reportan accuracies entre 88% y 95% con variantes de EfficientNet [13]. EfficientNet requiere su propio preprocesamiento interno mediante preprocess_input, diferente al estándar [0,1] de ResNet50.

C. Aumento de datos para balanceo de clases

El data augmentation es una técnica estándar para abordar el desbalance en datasets dermatoscópicos. Gessert et al. [7], ganadores de ISIC 2018, aplicaron volteos y transformaciones geométricas sobre

HAM10000. Buslaev et al. [8] documentaron en Albumentations que los parámetros estándar para imágenes médicas son: brightness y contrast $\pm 20\%$, saturación $\pm 20\%$ y rotaciones hasta 45 grados. En el presente trabajo se utilizaron parámetros más conservadores ($\pm 10\%$ brillo, contraste y saturación; $\pm 2\%$ tono; rotación ± 10 grados) para preservar las características diagnósticas de las lesiones. Iqbal et al. [10] validaron volteo horizontal, vertical y rotación de 90 grados sobre HAM10000 con ResNet.

D. Comparación de modelos en HAM10000

Un estudio comparativo entre ResNet50 y EfficientNet [15] encontró que EfficientNet superó a ResNet50 en datasets grandes, pero que en datasets de tamaño moderado los resultados son más variables y dependen de la estrategia de fine-tuning, esto teniendo en comparativa que la cantidad de imágenes del Dataset es de pequeña escala. Esteva et al. [16] demostraron que las CNN con Transfer Learning pueden alcanzar precisión comparable a dermatólogos certificados en clasificación de lesiones de piel.

V. METODOLOGÍA

La metodología se estructura en cuatro etapas: adquisición y balanceo del dataset, construcción de los modelos con distintas estrategias de Transfer Learning y evaluación comparativa de resultados.

A. Descripción del dataset HAM10000

El dataset HAM10000 fue publicado por Tschandl et al. [6] en 2018 y contiene 10,015 imágenes dermatoscópicas de siete tipos de lesiones. Para este trabajo se seleccionaron tres clases: nevo melanocítico (nv, n=6,705), melanoma (mel, n=1,113) y queratosis benigna (bkl, n=1,099), sumando 8,917 imágenes originales, siendo estas las que más volumen de imágenes tienen.

B. Preprocesamiento y balanceo de clases

Para resolver el problema de desbalance entre clases, primero se realizó undersampling sobre la clase nv y posteriormente se aplicó data augmentation a mel y bkl hasta alcanzar 3,000 imágenes por categoría. Las transformaciones utilizadas fueron seleccionadas buscando no alterar características importantes de las lesiones, siendo que:

Undersampling para nv: Se seleccionaron aleatoriamente 3,000 imágenes con semilla fija (random_state=42), para el data augmentation para mel y bkl se generaron imágenes sintéticas hasta completar 3,000 por clase usando Albumentations [8] con volteo horizontal ($p=0.5$) y vertical ($p=0.5$) válidos por ausencia de lateralidad anatómica [7, 10], así mismo una rotación de 90 grados ($p=0.5$); variación de color con brillo y contraste $\pm 10\%$, saturación $\pm 10\%$ y tono $\pm 2\%$ [8]; desenfoque gaussiano kernel 3×3 ($p=0.2$); y transformación affine con rotación ± 10 grados, escala $\pm 5\%$ y desplazamiento $\pm 5\%$ [7].

El dataset balanceado de 9,000 imágenes fue dividido en 70% entrenamiento (6,300), 15% validación (1,350) y 15% prueba (1,350) con división estratificada.

C. Configuración de los modelos

Los modelos se construyeron siguiendo el patrón de Transfer Learning de la guía oficial de TensorFlow [17], pero con estrategias distintas para cada arquitectura. ResNet50 se descargó desde Keras Applications [18] con 24,121,219 parámetros totales, de los cuales 24,064,003 son entrenables, esto se debe a que se aplicó Fine-Tuning completo dejando todas las capas de la red base descongeladas para que se adaptaran al dataset de lesiones de piel. EfficientNetB3 se descargó desde Keras Applications [19] con 11,183,922 parámetros totales y 3,951,485 entrenables, ya que en este caso se utilizó la estrategia de Extracción de Características: la red base de EfficientNetB3 se mantuvo congelada (trainable=False) y solo se entrenaron las capas del bloque de clasificación final. Esta decisión fue la que determinó la gran diferencia en parámetros entrenables entre ambos modelos.

D. Entrenamiento

Ambos modelos fueron entrenados con Adam [20] (learning rate= $1e-4$, batch size=32, máximo 30 épocas) en Google Colab con GPU NVIDIA T4. Se implementaron EarlyStopping (patience=7), ModelCheckpoint y ReduceLROnPlateau (factor=0.5, patience=3) de la API oficial de tf.keras [21]. ResNet50 completó las 30 épocas restaurando pesos de la época 28. EfficientNetB3 fue detenido por EarlyStopping en la época 23 restaurando pesos de la época 16.

VI. RESULTADOS

A. Desempeño de ResNet50

ReduceLROnPlateau se activó en las épocas 4, 14, 19 y 25, reduciendo el learning rate de $1e-4$ a $6.25e-6$. Las Tablas I y II presentan las métricas sobre el conjunto de prueba independiente de 1,350 imágenes.

TABLA I

Métricas por Clase en Conjunto de Prueba — ResNet50

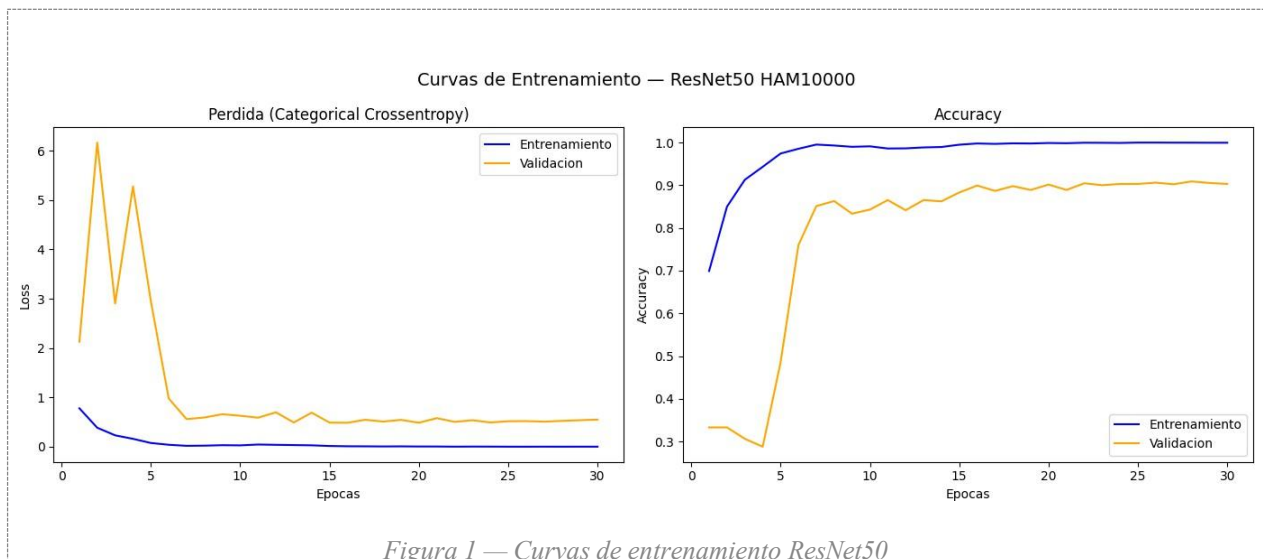
Clase	Precisión	Recall	F1-Score	Soporte
nv	0.92	0.90	0.91	450
mel	0.88	0.90	0.89	450
bkl	0.92	0.91	0.92	450
Macro avg	0.91	0.91	0.91	1,350

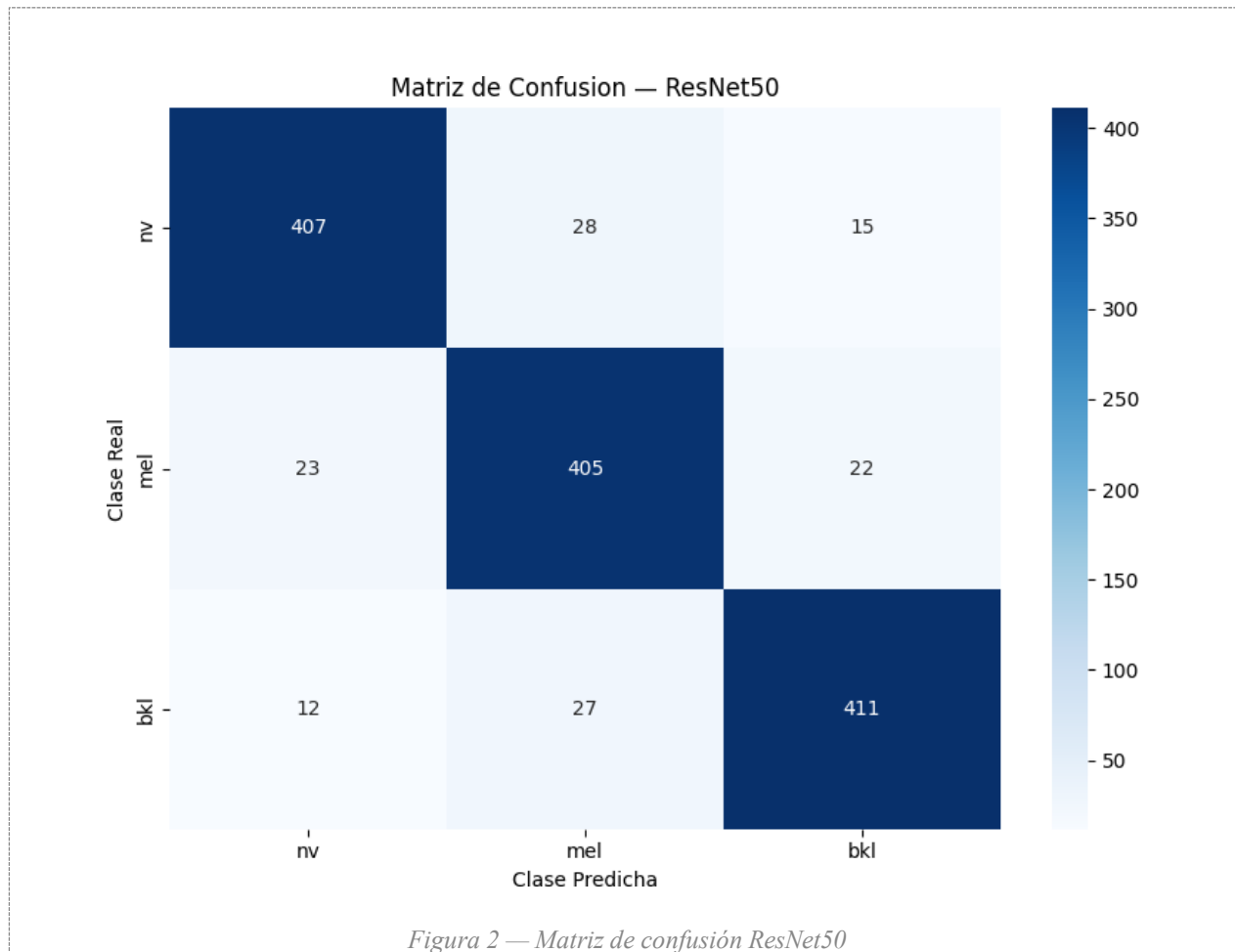
TABLA II

Métricas Globales en Conjunto de Prueba — ResNet50

Accuracy	Loss	F1-Score Macro	F1-Score Weighted
0.9059	0.4041	0.91	0.91

Las curvas de entrenamiento y la matriz de confusión de ResNet50 se presentan en las Figuras 1 y 2.





B. Desempeño de EfficientNetB3

ReduceLROnPlateau se activó en las épocas 14, 19 y 22. EarlyStopping detuvo el entrenamiento en la época 23 restaurando los pesos de la época 16 con val_accuracy=0.8770. Las Tablas III y IV presentan las métricas sobre el mismo conjunto de prueba en EfficientNet

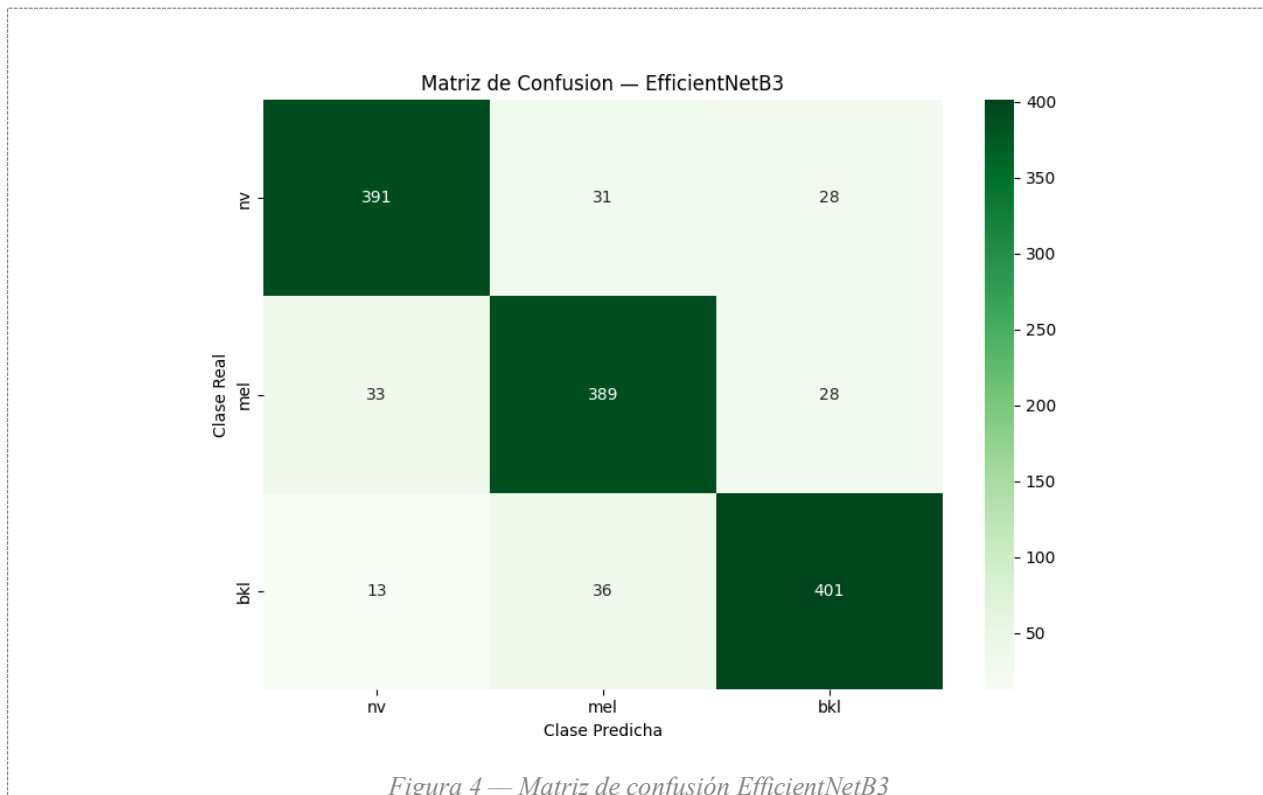
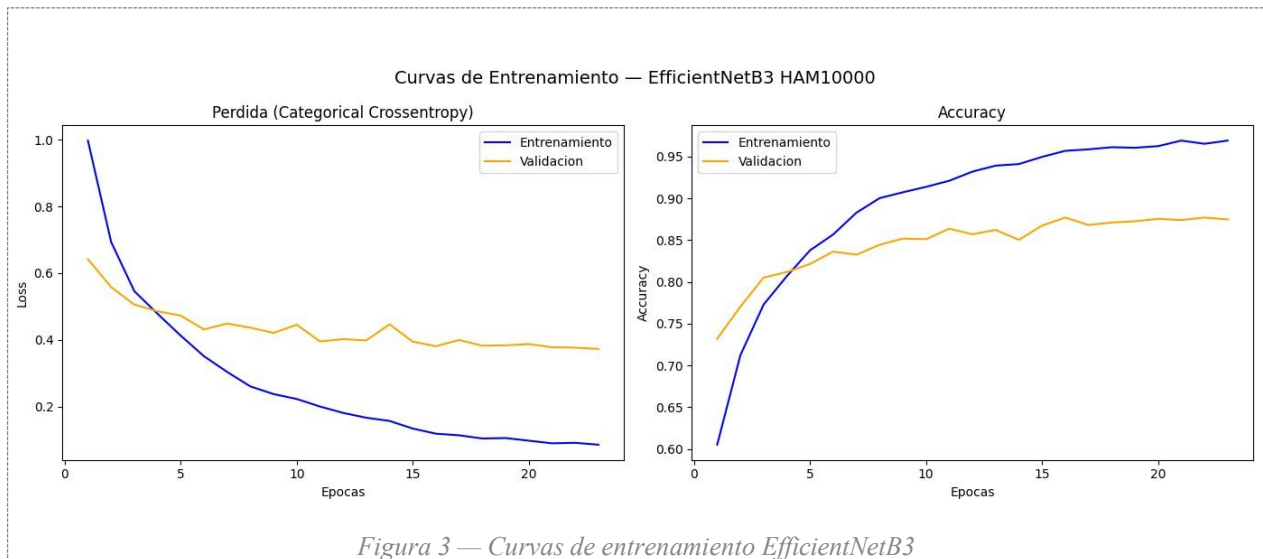
TABLA III
Métricas por Clase en Conjunto de Prueba — EfficientNetB3

Clase	Precisión	Recall	F1-Score	Soporte
nv	0.89	0.87	0.88	450
mel	0.85	0.86	0.86	450
bkl	0.88	0.89	0.88	450
Macro avg	0.88	0.87	0.87	1,350

TABLA IV
Métricas Globales en Conjunto de Prueba — EfficientNetB3

Accuracy	Loss	F1-Score Macro	F1-Score Weighted
0.8748	0.3620	0.87	0.88

Las curvas de entrenamiento y la matriz de confusión de EfficientNetB3 se presentan en las Figuras 3 y 4.



C. Comparación entre ambos modelos

La Tabla V presenta la comparación directa de ambos modelos bajo las mismas condiciones experimentales.

TABLA V
Comparación de Resultados — ResNet50 vs EfficientNetB3

Métrica	ResNet50	EfficientNetB3	Mejor
Accuracy	0.9059	0.8748	ResNet50
Loss	0.4041	0.3620	EfficientNetB3
F1-Score macro	0.91	0.87	ResNet50
Precisión nv	0.92	0.89	ResNet50
Precisión mel	0.88	0.85	ResNet50
Precisión bkl	0.92	0.88	ResNet50
Épocas entrenadas	30	23	EfficientNetB3
Parámetros totales	24,121,219	11,183,922	EfficientNetB3
Parámetros entrenables	24,064,003	3,951,485	EfficientNetB3

VII. DISCUSIÓN

Los resultados muestran que ResNet50 superó a EfficientNetB3 en accuracy (0.9059 vs 0.8748) y F1-score macro (0.91 vs 0.87). Aunque ambos modelos fueron evaluados con el mismo conjunto de prueba, las estrategias de entrenamiento fueron diferentes. En ResNet50 se permitió el reentrenamiento completo de la red, mientras que en EfficientNetB3 únicamente se entrenó la parte final del clasificador. Esto probablemente influyó de forma importante en los resultados obtenidos.

ResNet50 fue entrenado con Fine-Tuning completo, lo que significa que sus 24 millones de parámetros se reajustaron completamente al dataset de lesiones de piel, mientras que EfficientNetB3 se usó con Extracción de Características, manteniendo la red base congelada y entrenando únicamente el bloque final. Esto explica en gran medida la diferencia de resultados, ya que los patrones visuales de las lesiones dermatoscópicas son muy específicos y se benefician de que toda la red pueda adaptarse, siendo que la superioridad de EfficientNet reportada en algunos trabajos [14] suele darse cuando también se aplica Fine-Tuning, no cuando se congela la red base.

EfficientNetB3, por su parte, obtuvo un Loss menor (0.3620 vs 0.4041) y convergió más rápido (23 épocas vs 30), con considerablemente menos parámetros entrenables (3,951,485 vs 24,064,003), lo cual es completamente esperable dado que su red base estaba congelada. Eso representa una ventaja real en eficiencia computacional, y lo haría más conveniente en escenarios con recursos limitados o donde importe más la calibración de probabilidades que la accuracy cruda. Sin embargo, hay que tomar en cuenta que esta eficiencia viene de la mano de usar una estrategia más restrictiva de Transfer Learning.

Algo que vale la pena destacar es que ambos modelos tuvieron más dificultad clasificando melanoma (mel), con el F1-score más bajo de las tres clases (0.89 en ResNet50 y 0.86 en EfficientNetB3). Esto es clínicamente esperable: el melanoma comparte características visuales con nevos atípicos y queratosis seborreica pigmentada, y es una de las diferenciaciones más complicadas incluso para dermatólogos con experiencia [3].

Los parámetros de augmentation que usamos ($\pm 10\%$ brillo/contraste, $\pm 2\%$ tono, ± 10 grados rotación), más conservadores que los de Buslaev et al. [8], funcionaron bien para las tres clases sin distorsionar características diagnósticas relevantes.

Una limitación clara es el sobreajuste en ResNet50: la accuracy de entrenamiento llegó a superar el 99.9% en las últimas épocas, mientras que en prueba fue de 90.59%. Eso se explica por la cantidad de parámetros entrenables (24 millones) en comparación con el tamaño del dataset de entrenamiento (6,300 imágenes), una combinación que favorece el sobreajuste.

VIII. CONCLUSIONES

En este trabajo se comparó el desempeño de ResNet50 y EfficientNetB3 con Transfer Learning para clasificación automática de lesiones dermatoscópicas sobre el dataset HAM10000, siendo que ResNet50 resultó superior en accuracy (90.59% vs 87.48%) y F1-score macro (0.91 vs 0.87) sobre un dataset balanceado de 9,000 imágenes con tres clases diagnósticas.

La estrategia de balanceo con parámetros de augmentation justificados en literatura [7, 8, 10] demostró ser efectiva para producir modelos sin sesgo hacia la clase dominante, y EfficientNetB3 mostró ventajas claras en eficiencia como que, con menos parámetros, hay convergencia más rápida y menor loss, lo que lo convierte en una opción más práctica cuando los recursos computacionales son una limitante.

Por los resultados obtenidos, y al observarlos se pueden explicar principalmente la diferencia entre las CNN, siendo por la estrategia de Transfer Learning que se utilizó en cada modelo, ya que en la ResNet50 con Fine-Tuning completo pudo adaptar toda su arquitectura a las características específicas de las lesiones dermatoscópicas, mientras que EfficientNetB3 con Extracción de Características mantuvo su red base fija, lo que demuestra que la elección de la estrategia de entrenamiento puede tener un impacto igual o mayor que la elección de la arquitectura en sí.

Como trabajo futuro queda pendiente: (1) explorar augmentation más avanzado para las clases minoritarias; (2) evaluar el sistema con las 7 clases completas del HAM10000; (3) aplicar regularización adicional para reducir el sobreajuste en ResNet50 e (4) implementar un pipeline integrado de segmentación con UNet seguido de la clasificación.

REFERENCIAS

- [1] World Health Organization, "Skin cancers," WHO, 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/skin-cancers>
- [2] American Cancer Society, "Melanoma Skin Cancer Statistics," ACS, 2023. [Online]. Available: <https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html>
- [3] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *Lancet Oncology*, vol. 3, no. 3, pp. 159–165, 2002.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. CVPR*, 2016, pp. 770–778.
- [5] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. ICML*, 2019.
- [6] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, p. 180161, 2018.
- [7] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, R. Werner, and A. Schläpfer, "Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting," in *Proc. ISIC Skin Lesion Analysis Workshop, MICCAI*, 2018.
- [8] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.
- [9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [10] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, and J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," *Computerized Medical Imaging and Graphics*, vol. 88, 2021.
- [11] M. M. Musthafa, T. R. Mahesh, V. Vinoth Kumar, and S. Guluwadi, "Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification," *BMC Medical Imaging*, vol. 24, p. 207, 2024.
- [12] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *Journal of Biomedical Informatics*, vol. 86, pp. 25–32, 2018.
- [13] S. Shen, M. Xu, F. Zhang, P. Shao, H. Liu, L. Xu, C. Zhang, P. Liu, P. Yao, and R. X. Xu, "A low-cost high-performance data augmentation for deep learning-based skin lesion classification," *BME Frontiers*, vol. 2022, Art. no. 9765307, 2022.
- [14] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [15] J. Calderon, L. Fernandez, and M. Torres, "Skin Lesion Classification: A Transfer Learning Approach Using EfficientNets and ResNet50," in *Proc. International Conference on Biomedical Engineering*, 2020.
- [16] P. Tschandl et al., "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification," *npj Digital Medicine*, vol. 2, no. 1, p. 58, 2019.
- [17] TensorFlow, "Transfer learning and fine-tuning," TensorFlow Official Guide. [Online]. Available: https://www.tensorflow.org/tutorials/images/transfer_learning
- [18] Keras, "ResNet50 — Keras Applications," Official Documentation. [Online]. Available: <https://keras.io/api/applications/resnet/#resnet50-function>
- [19] Keras, "EfficientNetB3 — Keras Applications," Official Documentation. [Online]. Available: <https://keras.io/api/applications/efficientnet/#efficientnetb3-function>
- [20] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. ICLR*, 2015.
- [21] Keras, "Callbacks API," Official Documentation. [Online]. Available: <https://keras.io/api/callbacks/>