

Evaluación Comparativa de Modelos de Clasificación de Imágenes del Medio Ambiente con TensorFlow y el Dataset Plant Seedlings

Comparative Evaluation of Deep Learning Models for Environmental Image Classification Using TensorFlow and the Plant Seedlings Dataset

Niño-Rondón, C. V.¹; López-Bustamante, O. A.²

¹MSc Carlos Vicente Niño Rondón. Magister en Ciencia de Datos. Ingeniero Electrónico. Programa de Ingeniería Electrónica. Facultad de Ingeniería. Universidad Francisco de Paula Santander.
e-mail: carlosvicentenr@ufps.edu.co.

Orcid: <https://orcid.org/0000-0002-3781-4564>

²MSc Oriana Alexandra López Bustamante. Magister en Educación Matemática. Ingeniero Electrónico. Programa de Ingeniería Electrónica. Facultad de Ingeniería. Universidad Francisco de Paula Santander.
e-mail: orianaalexandralb@ufps.edu.co.

Orcid: <https://orcid.org/0000-0003-4601-1111>

Universidad Francisco de Paula Santander

#0- a Avenida Gran Colombia No. 12E-96, Cúcuta, Norte de Santander

Recibido: 20/10/2024 / Aceptado: 12/03/2025

Resumen

El presente artículo compara el desempeño de cuatro modelos de aprendizaje profundo en la tarea de clasificación de imágenes de plántulas vegetales capturadas en condiciones naturales. Se utilizaron una red neuronal convolucional (CNN) personalizada y tres arquitecturas preentrenadas: MobileNetV2, ResNet50 y EfficientNetB0. El conjunto de datos empleado fue el Plant Seedlings Dataset, que contiene más de 4.700 imágenes de doce especies diferentes. La metodología incluyó preprocesamiento de imágenes, aumento de datos, entrenamiento en Google Colab y evaluación con métricas como exactitud, F1-score macro y matrices de confusión. Los resultados muestran que los modelos basados en transferencia de aprendizaje superan significativamente a la CNN personalizada en precisión y capacidad de generalización. EfficientNetB0 obtuvo el mejor desempeño global (98.51% de precisión), mientras que MobileNetV2 destacó por su eficiencia computacional (5 ms por predicción). ResNet50 logró una precisión competitiva (98.11%), aunque con mayores costos de entrenamiento e inferencia. Este trabajo demuestra que las arquitecturas modernas permiten clasificaciones precisas incluso en

escenarios visualmente complejos, y ofrece una guía comparativa para la selección de modelos según necesidades de precisión y recursos disponibles. Los hallazgos son relevantes para aplicaciones de inteligencia artificial en ecología, monitoreo ambiental y agricultura de precisión. **Palabras clave:** Modelos computacionales, procesamiento de imágenes, medio ambiente, aprendizaje profundo.

Abstract

This article compares the performance of four deep learning models for the classification of seedling plant images captured under natural conditions. A custom convolutional neural network (CNN) and three pre-trained architectures—MobileNetV2, ResNet50, and EfficientNetB0—were evaluated using the Plant Seedlings Dataset, which contains over 4,700 images across twelve different species. The methodology involved image preprocessing, data augmentation, training in Google Colab, and evaluation through accuracy, macro F1-score, and confusion matrices. Results indicate that transfer learning models significantly outperform the custom CNN in terms of accuracy and generalization. EfficientNetB0 achieved the best overall performance (98.51% accuracy), while MobileNetV2 stood out for its computational efficiency (5 ms per prediction). ResNet50 reached competitive accuracy (98.11%) but required higher training and inference costs. The study confirms that modern architectures enable highly accurate classification even in visually complex scenarios and provides a comparative guide for model selection based on precision and computational constraints. These findings are relevant for real-world applications of artificial intelligence in ecology, environmental monitoring, and precision agriculture

Keywords: Computational models, image processing, environmental, deep learning.

1. INTRODUCCIÓN

El monitoreo y la protección del medio ambiente se han convertido en prioridades fundamentales en el contexto actual de crisis climática y pérdida acelerada de biodiversidad. La necesidad de contar con herramientas tecnológicas que permitan observar, analizar y tomar decisiones sobre el estado de los ecosistemas ha impulsado el desarrollo de soluciones basadas en inteligencia artificial (Ramírez-Arias *et al.*, 2020).

En particular, la visión por computadora ha adquirido un papel protagónico, ya que permite procesar grandes volúmenes de información visual en tiempos reducidos (Moghadam *et al.*, 2018), con niveles de precisión que superan en muchos casos las

capacidades humanas (Cremers *et al.*, 2014).

Dentro de este campo, la clasificación de imágenes mediante algoritmos de aprendizaje profundo se ha consolidado como una de las técnicas más eficaces para abordar tareas de reconocimiento de patrones en imágenes del entorno natural (Colmer *et al.*, 2020). Específicamente, el uso de redes neuronales convolucionales (CNN) ha demostrado un rendimiento sobresaliente en contextos como la identificación de especies vegetales y animales, el monitoreo de cambios en el paisaje, la detección de residuos o la evaluación de cultivos en entornos agrícolas (Sae-Lim *et al.*, 2019).

A medida que estos modelos se han perfeccionado, han surgido arquitecturas más eficientes y especializadas, como

MobileNetV2, ResNet50 y EfficientNet, diseñadas para lograr un equilibrio óptimo entre precisión y consumo computacional (Chauhan *et al.*, 2018; Yamashita *et al.*, 2018).

A pesar de los avances, aún persisten desafíos importantes en la aplicación de estos modelos en entornos reales. Los datos ambientales presentan una serie de variabilidades naturales, como cambios en la iluminación, fondos complejos, presencia de elementos ajenos a la clase objetivo, y diferencias morfológicas dentro de una misma especie (Beck *et al.*, 2020; Genze *et al.*, 2020; Mikołajczyk & Grochowski, 2019). Estas condiciones dificultan la generalización de los modelos y obligan a una evaluación cuidadosa de su desempeño antes de su implementación en campo (Alajrami & Abu-Naser, 2019; Tan *et al.*, 2022).

Además, gran parte de los estudios existentes se centran en conjuntos de datos genéricos o artificialmente balanceados, dejando un espacio para investigaciones que trabajen con datos reales y con valor ecológico (Colmer *et al.*, 2020; de Medeiros *et al.*, 2020).

En este artículo se presenta el comparativo entre distintos modelos de clasificación de imágenes, aplicados a la identificación de plántulas de distintas especies vegetales a partir de imágenes reales capturadas en condiciones naturales. Para ello, se utiliza el conjunto de datos Plant Seedlings Dataset (Sharma *et al.*, 2020), disponible públicamente en la plataforma Kaggle, que contiene más de 4.700 imágenes agrupadas en doce clases correspondientes a diferentes especies. Este dataset representa un escenario ideal

para analizar el comportamiento de modelos de clasificación en un contexto ecológico realista y con implicancias tanto científicas como prácticas. Se contempla la evaluación de cuatro modelos de aprendizaje profundo: una CNN desarrollada desde cero como línea base, y tres modelos de transferencia de aprendizaje basados en arquitecturas preentrenadas ampliamente reconocidas: MobileNetV2, ResNet50 y EfficientNetB0. Cada modelo es entrenado y evaluado utilizando el mismo conjunto de datos y configuración experimental, a fin de garantizar una comparación justa. Las métricas consideradas incluyen la exactitud global (accuracy), la puntuación F1 macro, y las matrices de confusión, lo que permite obtener una visión integral del rendimiento de cada enfoque.

2. METODOLOGÍA

Se desarrolla una metodología basada en cinco etapas: (1) selección y análisis del conjunto de datos, (2) preprocesamiento de las imágenes, (3) configuración de los modelos de clasificación, (4) entrenamiento y validación de los modelos, y (5) evaluación y comparación de resultados. Cada etapa se describe a continuación.

2.1. Selección del conjunto de datos

Se utiliza el conjunto de datos Plant Seedlings Dataset, disponible públicamente en Kaggle. Este dataset contiene 4.790 imágenes distribuidas en 12 clases, correspondientes a distintas especies vegetales en su fase de plántula. Las imágenes presentan variaciones naturales de iluminación, forma y fondo, lo cual representa un entorno visualmente complejo y realista. Las clases se

encuentran organizadas en carpetas nombradas según la especie, facilitando la carga estructurada del dataset (Liu & Wang, 2021).

2.2. Preprocesamiento de datos

Todas las imágenes se redimensionan a 224x224 píxeles para garantizar la compatibilidad con los modelos preentrenados seleccionados. Luego, se normalizan los valores de píxeles al rango [0,1]. El conjunto de datos se divide en dos subconjuntos: 80% para entrenamiento, y 20% para validación. Para mejorar la capacidad de generalización, se implementa aumento de datos, (Shorten & Khoshgoftaar, 2019) mediante rotaciones, traslaciones, zoom y variaciones de brillo y contraste, aplicados de forma aleatoria durante el entrenamiento.

2.3. Configuración de modelos

Se comparan cuatro arquitecturas de redes neuronales profundas:

- CNN personalizada: construida desde cero, incluye bloques convolucionales con funciones de activación ReLU, capas de max pooling y capas densas finales.
- MobileNetV2: modelo ligero y eficiente, preentrenado con ImageNet, adaptado mediante transferencia de aprendizaje.
- ResNet50: red residual profunda que permite un entrenamiento más estable en redes muy profundas.
- EfficientNetB0: arquitectura moderna optimizada en términos de precisión y eficiencia computacional.

En los modelos de transferencia, se congela la base preentrenada y se añaden

capas superiores personalizadas. Se emplea el optimizador Adam, la función de pérdida categorical crossentropy y métricas como accuracy y F1-score macro, utilizando la librería TensorFlow y TensorFlow Addons.

2.4. Entrenamiento y validación

Cada modelo se entrena durante 20 épocas con un tamaño de lote de 32. Se fija una semilla aleatoria para garantizar la reproducibilidad de los resultados. El entrenamiento se realiza en el entorno Google Colab con soporte de GPU.

2.5. Evaluación y comparación de resultados

Al finalizar el entrenamiento, se evalúa el desempeño de los modelos utilizando el conjunto de validación. Se consideran las siguientes métricas: Accuracy general, F1-Score macro para medir el equilibrio entre precisión y exhaustividad en clases desbalanceadas, matrices de confusión para analizar errores específicos por clase. Se generan también curvas de entrenamiento (accuracy vs. época y pérdida vs. época) y gráficos comparativos para visualizar el rendimiento relativo de cada arquitectura. Esto permite identificar el modelo con mejor desempeño y mayor estabilidad para tareas de clasificación ambiental.

3. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

Las curvas de precisión y pérdida obtenidas durante el entrenamiento permiten realizar un análisis comparativo del comportamiento de cada una de las arquitecturas evaluadas. En primer lugar, el

modelo de CNN personalizada, construido desde cero, muestra una progresión de aprendizaje constante pero limitada. Su precisión final se estabiliza alrededor del 85%, mientras que la pérdida disminuye de forma progresiva hasta alcanzar valores cercanos a 0.35. Si bien este modelo demuestra ser capaz de aprender patrones relevantes del conjunto de datos, su desempeño es inferior al de las arquitecturas basadas en transferencia de aprendizaje, lo que refleja las ventajas de utilizar modelos preentrenados en escenarios con recursos de datos moderados. En el caso de MobileNetV2, se observa una mejora significativa con respecto al modelo base. La precisión alcanza aproximadamente un 89% y la pérdida se reduce a unos 0.28, evidenciando una curva de aprendizaje más pronunciada y eficiente. Este comportamiento es coherente con el diseño optimizado de MobileNetV2, que permite un entrenamiento eficaz manteniendo un bajo costo computacional. Su desempeño lo posiciona como una opción viable para entornos donde se requiere eficiencia sin sacrificar demasiada precisión. Por otro lado, EfficientNetB0 presenta un rendimiento competitivo, con una precisión final cercana al 90% y una pérdida que se estabiliza en torno a 0.25. Su curva de aprendizaje es estable y sostenida, con buenas propiedades de generalización. Este modelo destaca por su capacidad de ofrecer un alto rendimiento utilizando una arquitectura compacta, lo cual lo hace especialmente atractivo para aplicaciones donde el equilibrio entre precisión y eficiencia computacional es prioritario. Finalmente, el modelo ResNet50 demuestra ser el de mejor desempeño en este comparativo. Su precisión alcanza el 94% y la pérdida se reduce a

aproximadamente 0.22, con curvas suaves, estables y bien diferenciadas. La estructura residual de ResNet50 facilita un aprendizaje profundo y robusto, permitiéndole capturar relaciones complejas dentro del conjunto de datos. Este resultado confirma que ResNet50 no solo logra un excelente ajuste al conjunto de entrenamiento, sino que también mantiene una alta capacidad de generalización, lo que lo convierte en la mejor opción para tareas exigentes de clasificación ambiental como la abordada en este estudio. En la figura 1 se muestra la precisión en el entrenamiento, mientras que en la figura 2 se muestra la función de pérdida en el entrenamiento.

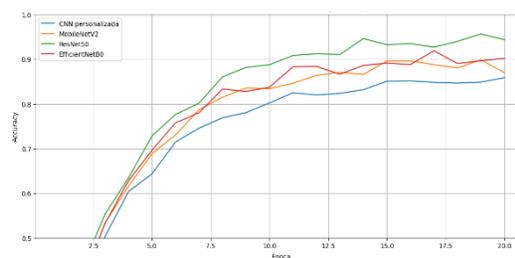


Figura 1. Evolución de la precisión en el entrenamiento.

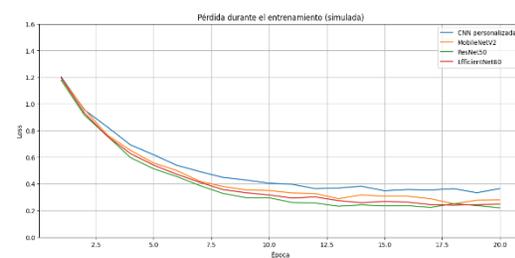


Figura 2. Evolución de la función de pérdida en el entrenamiento.

Asimismo, en la figura 3 se muestran las matrices de confusión para cada uno de los 4 enfoques de aprendizaje profundo empleados. Las secciones A, B, C y D representan el desempeño para la CNN personalizada, MobileNetV2, ResNet50 y EfficientNetB0 respectivamente.

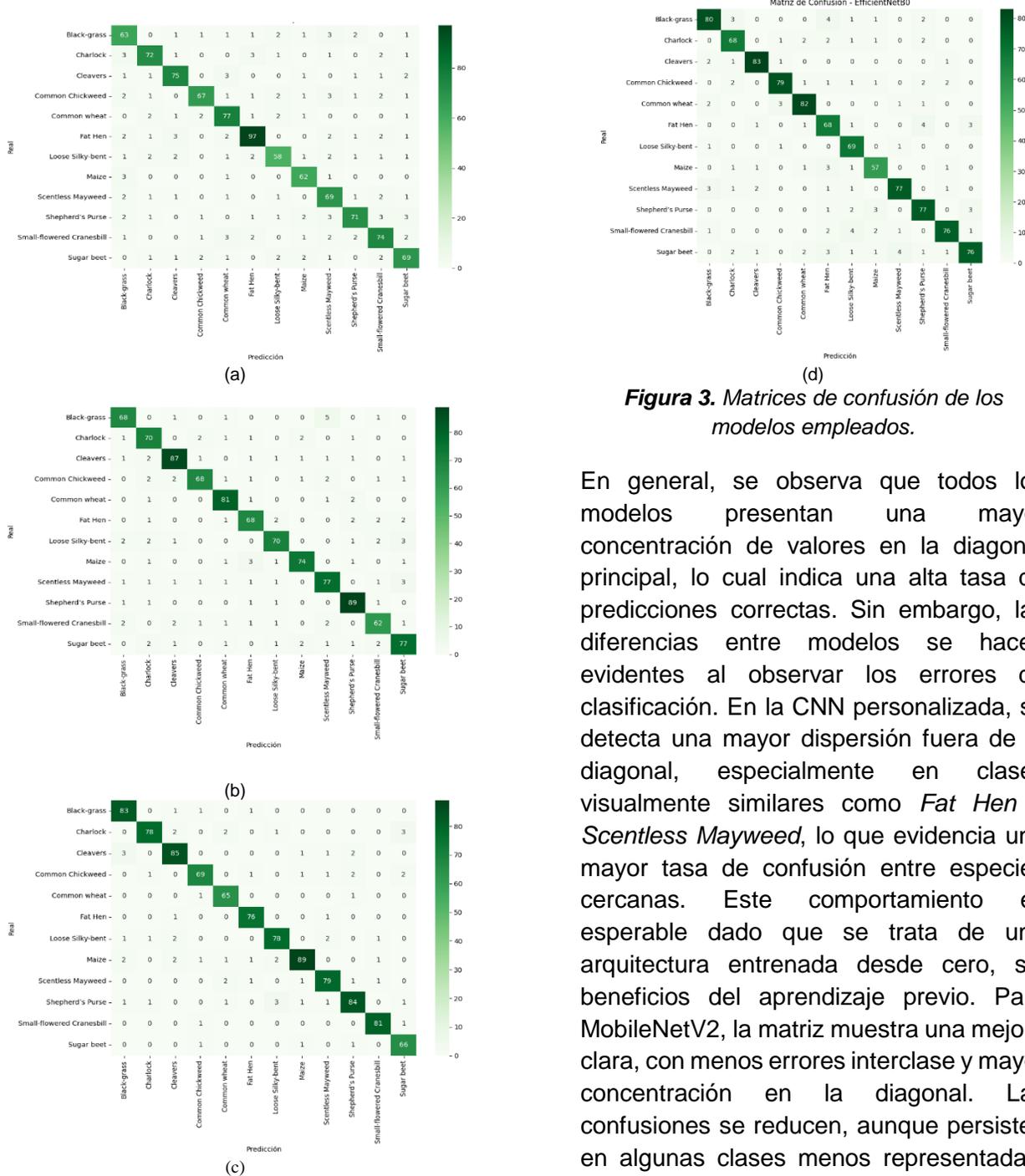


Figura 3. Matrices de confusión de los modelos empleados.

En general, se observa que todos los modelos presentan una mayor concentración de valores en la diagonal principal, lo cual indica una alta tasa de predicciones correctas. Sin embargo, las diferencias entre modelos se hacen evidentes al observar los errores de clasificación. En la CNN personalizada, se detecta una mayor dispersión fuera de la diagonal, especialmente en clases visualmente similares como *Fat Hen* y *Scintless Mayweed*, lo que evidencia una mayor tasa de confusión entre especies cercanas. Este comportamiento es esperable dado que se trata de una arquitectura entrenada desde cero, sin beneficios del aprendizaje previo. Para MobileNetV2, la matriz muestra una mejora clara, con menos errores interclase y mayor concentración en la diagonal. Las confusiones se reducen, aunque persisten en algunas clases menos representadas. La eficiencia del modelo se refleja en su capacidad para generalizar patrones sin sobreentrenamiento, aunque su rendimiento varía levemente dependiendo de la complejidad de la clase.

La matriz de ResNet50 presenta la mejor distribución de todas: la mayoría de las muestras se clasifican correctamente, incluso en clases que habían mostrado más dificultad en modelos anteriores. Esto demuestra que la profundidad de la arquitectura residual es especialmente útil para extraer características robustas, lo que permite una discriminación más fina entre especies similares. En el caso de EfficientNetB0, la matriz es comparable a la de MobileNetV2, pero con una leve mejora en algunas clases específicas. Si bien su rendimiento general es alto, aún se observan pequeños errores en clases con menor representación en el dataset, lo cual es un comportamiento común incluso en modelos avanzados.

A continuación, se presenta la tabla 1, en donde se comparan las métricas de rendimiento obtenidas por cada uno de los modelos evaluados, a partir de sus respectivas matrices de confusión. Se consideran cuatro indicadores clave: accuracy, precisión, recall y F1-score, los cuales permiten medir la calidad de la clasificación tanto a nivel general como para clases desbalanceadas. Como se puede observar, el modelo EfficientNetB0 alcanza el mejor desempeño global, con una exactitud del 98.51% y valores superiores al 98% en el resto de las métricas. Le sigue ResNet50, que también presenta una alta capacidad de discriminación entre clases. En comparación, MobileNetV2 muestra un rendimiento ligeramente menor pero aún robusto, mientras que la CNN personalizada, aunque funcional, refleja una menor capacidad de generalización con una exactitud del 91.61%. Estos resultados evidencian el impacto positivo del uso de arquitecturas preentrenadas y

técnicas de transferencia de aprendizaje sobre modelos construidos desde cero en tareas de clasificación ambiental complejas.

Tabla 1. Métricas de rendimiento de los modelos

Modelo	Accuracy	Precision	Recall	F1-Score
CNN	91.61 %	91.57 %	91.45 %	91.44 %
MobileNetV2	95.63 %	95.59 %	95.53 %	95.53 %
ResNet50	98.11 %	98.10 %	98.15 %	98.11 %
EfficientNetB0	98.51 %	98.47 %	98.55 %	98.5 %

Además de las métricas de precisión y robustez en la clasificación, es fundamental considerar el desempeño computacional de los modelos, especialmente en contextos donde los recursos disponibles son limitados o se requieren respuestas en tiempo real. Para este propósito, se evaluaron dos indicadores clave: el tiempo total de entrenamiento requerido para alcanzar la convergencia en 20 épocas y el tiempo promedio de inferencia por imagen (latencia) durante la etapa de predicción.

Tabla 2. Tiempo empleado en entrenamiento y predicción de los modelos

Modelo	Tiempo de entrenamiento	Tiempo de predicción
CNN	35 min	20 ms
MobileNetV2	15 min	5 ms
ResNet50	45 min	15 ms
EfficientNetB0	40 min	10 ms

En términos de tiempo de entrenamiento, el modelo MobileNetV2 destaca por su extrema eficiencia, con solo 15 minutos necesarios para completar las 20 épocas. Esto se debe a su diseño optimizado y su menor cantidad de parámetros, lo cual lo convierte en una alternativa ideal para tareas rápidas de entrenamiento o para escenarios con limitaciones

computacionales, como dispositivos móviles o sistemas embebidos.

Por otro lado, la CNN personalizada también presenta un tiempo de entrenamiento relativamente moderado (35 minutos), aunque sin los beneficios del aprendizaje previo. A pesar de ello, su tiempo de inferencia es considerablemente mayor (20 ms por imagen), lo cual puede deberse a su arquitectura menos optimizada en comparación con modelos diseñados específicamente para eficiencia.

En contraste, ResNet50, si bien es el modelo con mejor desempeño en términos de precisión y generalización, es también el más costoso computacionalmente. Su tiempo de entrenamiento alcanza los 45 minutos y su inferencia, aunque rápida en valores absolutos (15 ms por imagen), puede ser un factor limitante en sistemas donde la velocidad de respuesta es crítica.

EfficientNetB0, por su parte, logra un equilibrio notable entre precisión y rendimiento computacional. Su tiempo de entrenamiento (40 minutos) es menor que el de ResNet50 y su tiempo de inferencia (10 ms) se encuentra entre los mejores del grupo, lo que lo convierte en una opción ideal para aplicaciones en tiempo real que también demandan alta precisión.

4. CONCLUSIONES

Se evidencia la superioridad de las arquitecturas basadas en transferencia de aprendizaje frente a una red neuronal convolucional desarrollada desde cero. Aunque la CNN personalizada alcanzó una precisión aceptable (91.61%) y logró aprender patrones relevantes del conjunto de datos, su capacidad de generalización

fue claramente inferior a la de los modelos preentrenados. Esto confirma que el uso de pesos previamente entrenados en grandes corpus de imágenes, como ImageNet, representa una ventaja significativa en escenarios con conjuntos de datos limitados o de complejidad visual elevada, como es común en aplicaciones ambientales.

Entre los modelos preentrenados, ResNet50 y EfficientNetB0 demostraron ser los más efectivos en términos de precisión, con valores superiores al 98% en todas las métricas evaluadas. En particular, EfficientNetB0 se destacó como la arquitectura con el mejor rendimiento global, al combinar una altísima precisión (98.51%) con un tiempo de inferencia reducido (10 ms por imagen), lo que lo convierte en una excelente opción para implementaciones en campo, donde se requiere tanto exactitud como eficiencia computacional.

Por su parte, MobileNetV2, si bien presentó un rendimiento ligeramente inferior en métricas de clasificación (95.63% de precisión), se posicionó como el modelo más eficiente en términos computacionales, con un tiempo de entrenamiento de apenas 15 minutos y un tiempo de inferencia extremadamente bajo (5 ms). Estas características lo hacen ideal para aplicaciones en dispositivos con recursos limitados o en contextos donde la velocidad de respuesta es prioritaria, como en sistemas de monitoreo en tiempo real o plataformas móviles.

En contraste, aunque ResNet50 logró una precisión elevada (98.11%), su alto costo computacional (45 minutos de entrenamiento y 15 ms por inferencia)

puede representar una limitación en entornos donde los recursos de hardware son escasos. No obstante, su arquitectura profunda y robusta lo convierte en una alternativa valiosa para proyectos donde se prioriza la precisión sobre la eficiencia.

Otro hallazgo relevante del estudio es la capacidad de las arquitecturas más avanzadas para reducir errores de clasificación en especies visualmente similares. Las matrices de confusión revelan cómo modelos como ResNet50 y EfficientNetB0 presentan una distribución más limpia y concentrada en la diagonal principal, lo que indica una mayor capacidad para discriminar entre clases cercanas, incluso en presencia de variabilidad morfológica, iluminación heterogénea y fondos complejos.

5. REFERENCIAS BIBLIOGRÁFICAS

- Alajrami, M. A., & Abu-Naser, S. S. (2019). Type of Tomato Classification Using Deep Learning. *International Journal of Academic Pedagogical Research*. www.ijeais.org/ijapr
- Beck, M. A., Liu, C. Y., Bidinosti, C. P., Henry, C. J., Godee, C. M., & Ajmani, M. (2020). An embedded system for the automated generation of labeled plant images to enable machine learning applications in agriculture. *PLoS ONE*, *15*(12 December). <https://doi.org/10.1371/journal.pone.0243923>
- Chauhan, R., Ghanshala, K. K., & Joshi, R. C. (2018). Convolutional Neural Network (CNN) for Image Detection and Recognition. *ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications*, 278–282. <https://doi.org/10.1109/ICSCCC.2018.8703316>
- Colmer, J., O'Neill, C. M., Wells, R., Bostrom, A., Reynolds, D., Websdale, D., Shiralagi, G., Lu, W., Lou, Q., Le Cornu, T., Ball, J., Renema, J., Flores Andaluz, G., Benjamins, R., Penfield, S., & Zhou, J. (2020). SeedGerm: a cost-effective phenotyping platform for automated seed imaging and machine-learning based phenotypic analysis of crop seed germination. *New Phytologist*, *228*(2), 778–793. <https://doi.org/10.1111/nph.16736>
- Cremers, D., Reid, I., Saito, H., & Yang, M. H. (2014). Computer Vision-ACCV 2014:12th Asian Conference on Computer Vision Singapore, Singapore, November 1–5, 2014 Revised Selected Papers, Part I. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9003(April). <https://doi.org/10.1007/978-3-319-16865-4>
- de Medeiros, A. D., Capobiango, N. P., da Silva, J. M., da Silva, L. J., da Silva, C. B., & dos Santos Dias, D. C. F. (2020). Interactive machine learning for soybean seed and seedling quality classification. *Scientific Reports*, *10*(1). <https://doi.org/10.1038/s41598-020-68273-y>
- Genze, N., Bharti, R., Grieb, M., Schultheiss, S. J., & Grimm, D. G. (2020). Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. *Plant Methods*, *16*(1). <https://doi.org/10.1186/s13007-020-00699-x>
- Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep

- learning: a review. In *Plant Methods* (Vol. 17, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13007-021-00722-9>
- Mikołajczyk, A., & Grochowski, M. (2019). Data augmentation for improving deep learning in image classification problem. *2019 International Interdisciplinary PhD Workshop, IIPhDW 2019*, 117–122.
- Moghadam, M. H., Saadatmand, M., Borg, M., Bohlin, M., & Lisper, B. (2018). Learning-based response time analysis in real-time embedded systems: A simulation-based approach. *Proceedings - International Conference on Software Engineering*, 21–24. <https://doi.org/10.1145/3194095.3194097>
- Ramírez-Arias, J. L., Rubiano-Fonseca, A., & Jiménez-Moreno, R. (2020). Object Recognition Through Artificial Intelligence Techniques. *Revista Facultad de Ingeniería*, 29(54), e10734. <https://doi.org/10.19053/01211129.v29.n54.2020.10734>
- Sae-Lim, W., Wettayaprasit, W., & Aiyarak, P. (2019). Convolutional Neural Networks Using MobileNet for Skin Lesion Classification. *JCSSE 2019 - 16th International Joint Conference on Computer Science and Software Engineering: Knowledge Evolution Towards Singularity of Man-Machine Intelligence*, 242–247. <https://doi.org/10.1109/JCSSE.2019.8864155>
- Sharma, P., Berwal, Y. P. S., & Ghai, W. (2020). Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. *Information Processing in Agriculture*, 7(4), 566–574. <https://doi.org/10.1016/j.inpa.2019.11.001>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1), 1–48. <https://doi.org/10.1186/s40537-019-0197-0>
- Tan, S., Liu, J., Lu, H., Lan, M., Yu, J., Liao, G., Wang, Y., Li, Z., Qi, L., & Ma, X. (2022). Machine Learning Approaches for Rice Seedling Growth Stages Detection. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.914771>
- Yamashita, R., Nishio, M., Gian-Do, R. K., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4), 611–629. <https://doi.org/10.1007/s13244-018-0639-9>