



Ciencia Latina
Internacional

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), noviembre-diciembre 2024,
Volumen 8, Número 6.

https://doi.org/10.37811/cl_rcm.v8i6

ANÁLISIS DE DESEMPEÑO DE DIFERENTES ARQUITECTURAS DE REDES NEURONALES CONVOLUCIONALES (CNN) PARA CLASIFICACIÓN AUTOMÁTICA DE DESECHOS SÓLIDOS

**PERFORMANCE ANALYSIS OF DIFFERENT
CONVOLUTIONAL NEURAL NETWORK (CNN)
ARCHITECTURES FOR AUTOMATIC SOLID
WASTE CLASSIFICATION**

Jhonny Darwin Ortiz-Mata

Universidad Estatal de Milagro, Ecuador

Xiomara Jael Oleas Velez

Universidad Estatal de Milagro, Ecuador

David Elías Dager López

Universidad Estatal de Milagro, Ecuador

Marcia Maribel Moyolema Amboya

Universidad Estatal de Milagro, Ecuador

DOI: https://doi.org/10.37811/cl_rcm.v8i6.15455

Análisis de Desempeño de Diferentes Arquitecturas de Redes Neuronales Convolucionales (CNN) para Clasificación Automática de Desechos Sólidos

Jhonny Darwin Ortiz Mata¹jortizm2@unemi.edu.ec<https://orcid.org/0000-0003-0466-4093>Universidad Estatal de Milagro
Ecuador**Xiomara Jael Oleas Velez**xoleasv@unemi.edu.ec<https://orcid.org/0009-0009-7421-8452>Universidad Estatal de Milagro
Ecuador**David Elías Dager López**ddagel@unemi.edu.ec<https://orcid.org/0000-0001-6663-6149>Universidad Estatal de Milagro
Ecuador**Marcia Maribel Moyolema Amboya**mmoyolemaa@unemi.edu.ec<https://orcid.org/0009-0003-6687-9409>Universidad Estatal de Milagro
Ecuador

RESUMEN

El rápido aumento en la generación de residuos sólidos ha dado lugar a gigantescos desafíos medioambientales, algunas de las razones son: el crecimiento de la población, el consumo masivo y las prácticas poco adecuadas de gestión de residuos. Es necesario abordar estos problemas para lograr un desarrollo sostenible y una economía circular. Los métodos tradicionales de gestión de residuos han quedado obsoletos y se necesitan soluciones innovadoras que incorporen tecnologías con inteligencia artificial. Este trabajo describe la clasificación automática de residuos sólidos basada en arquitecturas CNN con alta precisión y eficiencia en la clasificación en tiempo real. Se evalúan varias arquitecturas de CNN, VGG16-BN, ResNet50, Inception V3, Xception y MobileNetV3 Large, entre los materiales de desechos están plástico, vidrio, metal, cartón y papel. Los resultados muestran que MobileNetV3 Large FT es la arquitectura más eficaz, alcanzando la mayor precisión con un 96.89% y la menor pérdida con un 10.87%, destacando tanto en precisión de clasificación como en eficiencia computacional. Este trabajo pone de relieve hasta qué punto los modelos de IA basados en CNN pueden contribuir a mejorar la clasificación de residuos, reduciendo los errores humanos y ayudando potencialmente a mantener la sostenibilidad medioambiental.

Palabras clave: desechos sólidos, CNN, xception, fine-tuning, mobilenetl

¹ Autor principal.

Correspondencia: jortizm2@unemi.edu.ec

Performance Analysis of Different Convolutional Neural Network (CNN) Architectures for Automatic Solid Waste Classification

ABSTRACT

The rapid increase in solid waste generation has led to significant environmental challenges. Some of the contributing factors include population growth, mass consumption, and inadequate waste management practices. Addressing these issues is essential to achieve sustainable development and a circular economy. Traditional waste management methods have become obsolete, necessitating innovative solutions that incorporate artificial intelligence technologies. This study focuses on the automatic classification of solid waste using CNN architectures, achieving high precision and real-time classification efficiency. Several CNN architectures are evaluated, including VGG16-BN, ResNet50, Inception V3, Xception, and MobileNetV3 Large. The waste materials analyzed include plastic, glass, metal, cardboard, and paper. Results indicate that MobileNetV3 Large FT is the most effective architecture, achieving the highest accuracy at 96.89% and the lowest loss at 10.87%, excelling in both classification precision and computational efficiency. This study highlights the extent to which CNN-based AI models can improve waste classification, reduce human errors, and potentially contribute to environmental sustainability.

Keywords: solid waste, CNN, xception, fine-tuning, mobilenetl

Artículo recibido 15 octubre 2024

Aceptado para publicación: 20 noviembre 2024



INTRODUCCIÓN

El aumento exponencial de los residuos sólidos en las últimas décadas ha generado graves problemas medioambientales, impulsados por factores como el crecimiento demográfico, el consumo masivo de productos y la limitada conciencia sobre la gestión adecuada de desechos. Esta situación ha intensificado la crisis de contaminación, exigiendo la implementación de soluciones tecnológicas innovadoras y sostenibles alineadas con los Objetivos de Desarrollo Sostenible (ODS) y los principios de las 9R (Narayan, 2021).

En este contexto, se resalta la importancia de los enfoques basados en inteligencia artificial (IA) para abordar estos desafíos medioambientales. Cabe destacar que, las redes neuronales convolucionales (CNN) representan una herramienta eficaz para la clasificación automática de residuos mediante el análisis de imágenes (White et al., 2020). Este trabajo propone el desarrollo de una aplicación de CNN, diseñada para clasificar residuos sólidos de manera precisa, con un enfoque en materiales reciclables como plástico, vidrio, metal, cartón y papel. Además, se integran técnicas avanzadas de ajuste fino y aumento de datos para optimizar el desempeño del modelo, especialmente en entornos con recursos computacionales limitados.

Revisión del Estado del Arte

La clasificación automática de residuos ha evolucionado significativamente con el uso de arquitecturas CNN. Modelos como VGG16-BN, ResNet50, Inception V3, Xception y MobileNetV3-Large han sido evaluados extensamente en términos de precisión, pérdida y eficiencia computacional. Por ejemplo, VGG16-BN ha demostrado una precisión del 93.37% en la clasificación de residuos orgánicos y reciclables tras solo 5 épocas de entrenamiento (Da et al., 2023). Asimismo, ResNet50 se destaca por su arquitectura residual, que permite entrenar redes profundas sin pérdida de precisión, alcanzando una eficacia del 91.8% con solo 7 épocas de entrenamiento (He et al., 2015).

En aplicaciones móviles y entornos con recursos computacionales limitados, modelos ligeros como MobileNetV3-Large han demostrado ser ideales. Este modelo, con una precisión aproximada del 90%, ha mostrado eficiencia tanto en términos de rendimiento como de adaptabilidad a entornos industriales (Sayed et al., 2024).

Por otro lado, arquitecturas como Xception y Inception V3 destacan en escenarios complejos, permitiendo clasificaciones precisas de materiales reciclables con características similares, como vidrio, plástico y metal (Bishop et al., 2023; Sadeghnezhad & Salem, 2024).

La elección de la arquitectura adecuada depende no solo de la precisión, sino también de las restricciones de almacenamiento y velocidad. Estudios recientes han explorado técnicas de optimización como la poda de parámetros redundantes para reducir el tamaño de los modelos sin pérdida significativa de precisión (Brockmann & Schlippe, 2024). Sin embargo, aún existen desafíos relacionados con la generalización de los modelos en entornos reales y con la evaluación de métricas clave como el coeficiente de correlación de Matthews (MCC), que permite medir el rendimiento en conjuntos de datos desequilibrados (Chollet, 2021).

En síntesis, las investigaciones previas destacan el potencial de las arquitecturas CNN para mejorar la gestión de residuos sólidos. Este estudio busca avanzar en esta línea al evaluar el rendimiento de varias arquitecturas preentrenadas y optimizadas, identificando la más adecuada para un sistema eficiente y accesible.

Planteamiento del Problema

La administración ineficaz de desechos sólidos ha incrementado las inquietudes medioambientales y los desafíos de salud pública, transformándose en uno de los principales impedimentos para llevar a cabo los Objetivos de Desarrollo Sostenible de la ONU. Con el incremento en los niveles de consumo y la ausencia de sistemas apropiados para el manejo de desechos, las oportunidades de un reciclaje eficaz de los materiales se vuelven cada vez menos factibles. Pese a que hay ciertas iniciativas enfocadas en optimizar el reciclaje, los retos tecnológicos y operativos, obstaculizan la categorización exacta y eficaz de los materiales reciclables.

En este escenario, las soluciones fundamentadas en inteligencia artificial poseen un enorme potencial, en particular las redes neuronales convolucionales, que pueden ser utilizadas en el análisis de imágenes y realizar la categorización automática de desechos. No obstante, la ausencia de un análisis comparativo entre diversas arquitecturas ya entrenadas como Xception, MobileNetV3-Large, Inception V3, DenseNet121, ResNet50, VGG16-BN y YOLOv5L dificulta la creación de sistemas más sólidos, capaces de ajustarse a diversos contextos.

Realizar una investigación crítica en la evaluación de estas arquitecturas es esencial con respecto a su eficacia en términos de precisión, pérdida, recall, F1-score y coeficiente de correlación de Matthews. Este tipo de evaluación detallada es esencial para facilitar la identificación de las arquitecturas de mejor rendimiento; también ayudaría a enmarcar métodos de optimización que aseguren un mejor desempeño acorde con la realidad. Diseño de sistemas correctos y fáciles de adquirir.

Justificación y Relevancia

La metodología de este trabajo se centra en la evaluación de arquitecturas CNN previamente entrenadas. Todas las versiones de Xception, MobileNetV3-Large, Inception V3, DenseNet121, ResNet 50, VGG16-BN y YOLOv5L fueron mejoradas descongelando las capas más bajas e incorporando un bloque secuencial personalizado para ajustar la salida a una categorización en 30 categorías (Kunwar, 2023).

La importancia de este trabajo experimental es la necesidad de modelos entrenados para la problemática de clasificación de desechos, asegurando la eficiencia del rendimiento de las arquitecturas mediante un sistema de métricas cuantitativas.

Objetivo General

- Determinar la arquitectura CNN más adecuada para la clasificación de residuos sólidos mediante el análisis de su rendimiento en términos de precisión, eficiencia computacional y capacidad de generalización.

Objetivos Específicos

1. Potenciar los modelos seleccionados mediante técnicas avanzadas de ajuste fino, la adición de bloques secuenciales personalizados; con el fin de aumentar la generalización de la arquitectura en diferentes entornos (Kunwar, 2023).
2. Analizar y comparar la eficiencia computacional y la precisión de diferentes arquitecturas CNN en los conjuntos de prueba y validación, mediante de métricas de exactitud y perdida.
3. Evaluar el rendimiento de diferentes CNNs preentrenados para la clasificación automática de residuos sólidos, utilizando indicadores de rendimiento clave como la precisión, la pérdida, la recall, la puntuación F1 y el coeficiente de correlación de Matthews (Da et al., 2023).

METODOLOGÍA

Se trata de un estudio cuantitativo incluido en la categoría de investigación experimental. Consiste en la manipulación de diferentes arquitecturas de Redes Neuronales Convolucionales para la resolución del problema de clasificación de residuos sólidos. El estudio también utilizará un marco de evaluación basado en indicadores cuantitativos: precisión, pérdida en los conjuntos de validación y prueba; con la finalidad de evaluar y justificar el rendimiento de los modelos presentados.

El primer paso en este flujo de trabajo de investigación sobre clasificación de residuos es la adquisición de datos. El conjunto de datos se compone de imágenes de residuos reciclables y domésticos obtenidas de la plataforma Kaggle, y se carga en Google Colab, para aprovechar sus recursos computacionales que ofrece la plataforma, para el entrenamiento del modelo. El preprocesamiento de datos incluye la normalización y el aumento. La normalización mantiene los valores de los píxeles en un rango uniforme para facilitar el aprendizaje y optimizar el rendimiento del modelo, mientras que el aumento de datos aplica transformaciones como la rotación, los ajustes de brillo y la inversión para aumentar la diversidad de la muestra.

La base de datos se divide en tres conjuntos: 60% entrenamiento, 20% validación y 20% pruebas. Esta división garantiza un aprendizaje eficaz y una evaluación objetiva del rendimiento del modelo en cada etapa. El entrenamiento se realiza utilizando la función `CrossEntropyLoss` junto con el optimizador Adam; esta decisión se fundamenta en el diseño arquitectónico de cada CNN en PyTorch, la capa de salida corresponde a tipo lineal, por lo tanto, la función de pérdida `CrossEntropyLoss` incluye la funcionalidad de Softmax de manera implícita; por otro lado, se usa Adam dado que combina las mejores características de RMSProp y Momentum, lo que da como resultado el equilibrio óptimo entre velocidad de convergencia y estabilidad durante el entrenamiento. (Elgendy, 2020). El entrenamiento se ejecutó lo largo de 20 épocas, los parámetros del modelo se ajustan iterativamente para maximizar la precisión de la clasificación.

Al final de cada época, se guardan los mejores pesos del modelo, también se debe evitar el sobreajuste y asegurar la eficacia del modelo.

El modelo se evalúa utilizando métricas visuales como la precisión y las matrices de confusión para identificar las clases en la que el modelo clasifica correctamente (Stevens et al., 2020).

Adoptando este minucioso proceso, el modelo se prepara para su integración en una aplicación móvil que incorpora sistemáticamente el preprocesamiento, el entrenamiento y la evaluación para la clasificación de residuos

Procedimiento

Tabla 1. Categorías de la base de datos

Código	Clase	Código	Clase
c0	Latas de aerosol	c15	Papel de oficina
c1	Latas de aluminio para alimentos	c16	Vasos de papel
c2	Latas de refresco de aluminio	c17	Tapas de plástico
c3	Cajas de cartón	c18	Botellas de plástico de detergente
c4	Envases de cartón	c19	Envases de plástico para alimentos
c5	Ropa	c20	Bolsas de plástico
c6	Posos de café	c21	Botellas de plástico de refrescos
c7	Cubiertos de plástico desechables	c22	Pajitas de plástico
c8	Cáscaras de huevo	c23	Bolsas de basura de plástico
c9	Residuos de alimentos	c24	Botellas de agua de plástico
c10	Botellas de vidrio para bebidas	c25	Zapatos
c11	Envases de vidrio para bebidas	c26	Latas de acero
c12	Tarros de vidrio para alimentos	c27	Vasos de poliestireno
c13	Revistas	c28	Envases de comida de poliestireno
c14	Periódicos	c29	Bolsas de té

En la Tabla 1 se presenta un resumen de las 30 diferentes clases del conjunto de datos de clasificación de residuos reciclables y domésticos. Cada clase se ha codificado de c0 a c29 para garantizar un etiquetado ordenado en el conjunto de datos. La Tabla 1 muestra la representación en el conjunto de datos de la clasificación de residuos en el mundo real. De este modo, se mejora la generalidad del modelo en distintos tipos de residuos y se obtiene un rendimiento más fiable.

El conjunto de datos dividido en tres subconjuntos: entrenamiento, validación y prueba. En esta configuración, las imágenes se dividen en tres subconjuntos: entrenamiento (60% o 36.000 imágenes), y conjuntos de validación y prueba con un 20% cada uno (12.000 imágenes). Gran parte de la base de datos se establece para el conjunto de entrenamiento, esto se debe para que el modelo tenga suficiente cantidad de datos para aprender patrones complejos y generalizar mejor con nuevos datos.

Esta distribución equitativa proporciona un equilibrio adecuado para que los subconjuntos de validación y prueba midan el rendimiento del modelo. El conjunto de validación se utiliza en primer lugar para ajustar los hiperparámetros y evitar el sobreajuste. Después, siguiendo la sugerencia de (Atienza, 2020), el conjunto de prueba proporciona una evaluación objetiva del rendimiento en datos no vistos. De este modo, la distribución constituye una base sólida para el entrenamiento, la validación efectiva y la evaluación imparcial del modelo.

Preprocesamiento de datos

Las transformaciones ligeras que se aplican buscan aumentar la cantidad de datos disponible para el procesamiento de imágenes en las arquitecturas. La clase `WasteDataset` es la responsable de cargar y aplicar las transformaciones antes de alimentarlas al modelo con cada arquitectura. La primera transformación es la Rotación, genera variaciones angulares para simular diferentes perspectivas, la rotación aplicada a cada entrada ronda en el intervalo $[-20, 20]$.

El ajuste de brillo es la segunda transformación aplicada, el factor $B(\alpha)$ establece un valor aleatorio entre $[0.7, 1.3]$. Este ajuste permite que el modelo se adapte a diferentes condiciones de iluminación (Koul et al., 2019). La cuarta transformación corresponde a la técnica de espejado horizontal reduce la sensibilidad del modelo a la orientación del objeto. Esta técnica ejecuta una versión espejada de la imagen. El factor filtro gaussiano, la quinta transformación, establece $G(\sigma)$ en un radio, donde σ es un número aleatorio en el rango $[0.1, 2.0]$, su función es suavizar, difuminando detalles y reduciendo el ruido según el valor radio aleatorio.

La sexta técnica de aumento de datos es la normalización estandariza los valores de los píxeles, los delimita a un rango $[-1, 1]$ según la media y desviación estándar de cada canal de color. Para el procesamiento de imágenes, se aplicó una media de 0.5 y una desviación estándar de 0.5 en cada canal RGB. Según (Dawani, 2020) resalta la importancia de esta transformación para la consistencia y precisión durante el entrenamiento del modelo.

La transformación `RandomAffine` se encarga de realizar modificaciones aleatorias a la imagen objetivo, en la incluye rotaciones, traslaciones y escalados, esto mejora la capacidad del modelo para reconocer objetos en diferentes posiciones y tamaños, según Morales (Morales M., 2020).

Arquitecturas CNN

Tabla 2: Arquitecturas de modelos CNN.

Arquitectura	Parámetros Totales	Parámetros Entrenables	Precisión de Prueba (%)	Pérdida (%)	Tamaño del Modelo (MB)	Principales Ventajas
MobileNetV3 Large FT	8,171,526	1,080,350	96.89	11.45	219.17	Eficiente, ligera y con alto rendimiento en móviles.
Xception FT	45,875,270	25,068,318	96.41	15.56	698.12	Alta precisión y capacidad de aprendizaje profundo.
DenseNet121 FT	8,543,646	1,591,838	96.24	14.88	926.77	Conexiones densas que optimizan la reutilización.
Inception V3 FT	25,045,280	2,638,366	86.52	43.94	358.23	Balance entre profundidad y eficiencia.
ResNet50 FT	25,636,214	2,638,366	94.07	31.80	465.27	Robusta para clasificación en lotes grandes.
VGG16-BN FT	139,004,510	21,516,830	96.54	15.25	4,035.07	Precisión en datos preentrenados, pero pesada.

La tabla 2 destaca que MobileNetV3 Large FT es la arquitectura más eficiente para la clasificación de residuos, alcanzando la mayor precisión de prueba (96,49%) y la menor pérdida (11,45%), con un tamaño de modelo relativamente pequeño (219.17 MB). Estas características la convierten en una opción ideal para aplicaciones móviles, donde la capacidad de procesamiento y almacenamiento es limitada. Además, la relación entre parámetros totales y entrenables en MobileNetV3 Large FT demuestra su equilibrio entre eficiencia computacional y adaptabilidad, optimizando su desempeño en dispositivos de bajo consumo energético. Para todos los modelos CNN que se muestran en la Tabla 2, se utilizó la técnica de ajuste “*Fine-Tuning*” (FT) para mejorar la precisión.

En contraste, VGG16-BN FT presenta un alto número de parámetros totales y un tamaño considerable (4,035.07 MB), lo que dificulta su implementación en entornos móviles, a pesar de su buena precisión de prueba (94.55%). La comparación entre las arquitecturas evidencia que la elección del modelo depende no solo de la precisión sino también de las restricciones de almacenamiento y velocidad, donde MobileNetV3 Large FT sobresale como la solución más equilibrada.

Hiperparámetros

Tabla 3: Parámetros de configuración

Hiperparámetros	Valores		
Tasa de aprendizaje	1×10^{-4}	5×10^{-4}	1×10^{-5}
Batch Size	32	16	
Optimizer	Adam	Weight decay	1×10^{-4}
Función de Perdida	CrossEntropyLoss		
Épocas	20	50	30

La Tabla 3 presenta los principales parámetros utilizados en el entrenamiento de los siete modelos de redes neuronales convolucionales. Cada intento presenta una configuración específica para la tarea de clasificación de residuos, en la que la tasa de aprendizaje varía según las métricas de precisión y pérdida. La tasa de aprendizaje se controla mediante código de programación. De este modo, se llevará a cabo una reducción de la tasa de aprendizaje en un factor de 0,05 si no se produce ninguna mejora en la métrica de pérdidas a lo largo de cinco épocas consecutivas, optimizando la adaptación del modelo a las nuevas condiciones (Bianco et al., 2018).

(Bischi et al., 2021), expresa que la tasa de aprendizaje es uno de los hiperparámetros más importantes, define la magnitud de la actualización de pesos e influye directamente en la estabilidad del modelo y la velocidad de convergencia. Mientras tanto, la regularización L2, comúnmente conocida como caída de pesos, penaliza los pesos grandes y fomenta la generalización adecuada, una estrategia sugerida por (Novello et al., 2023) para evitar el sobreajuste. El escalado dinámico del gradiente con “GradScaler” optimiza la eficiencia de la memoria y el rendimiento durante el entrenamiento de precisión mixta (Ilievski et al., 2017).

Métricas CNN

El sistema de métricas para la evaluación del rendimiento de los modelos CNN, permite comprender la eficacia y comportamiento de la arquitectura. Entre las métricas se encuentra la precisión, que, durante el entrenamiento, se divide en “precisión de entrenamiento” y “precisión de validación”. Esta medida es la relación entre predicciones correctas y predicciones totales, lo que ayuda a evaluar cómo el modelo generaliza a nuevos conjuntos de datos diferentes (Elgendy, 2020).

Del mismo modo, la métrica de pérdida indica la proporción de predicciones erróneas con respecto al número total de predicciones. Además, al igual que la métrica anterior, se clasifica en “pérdida de entrenamiento” y “pérdida de validación”, lo que da una idea aproximada de los errores en las fases de entrenamiento y validación. La función de pérdida se convierte en una parte crucial, ya que permite ajustar los parámetros del modelo para reducir los errores y aumentar el rendimiento (Ayyadevara & Reddy, 2020).

La métrica “*Precision*” mide la proporción de verdaderos positivos de todas las instancias clasificadas como positivas; evalúa la precisión en la identificación correcta de la clase positiva. Esto resulta útil cuando un modelo debe clasificar un gran número de falsos positivos.

De forma complementaria, “*recall*” indica la fracción de verdaderos positivos entre todas las instancias que deberían haber sido clasificadas como positivas, una medida de la capacidad del modelo para identificar correctamente todos los casos positivos esperados (Stevens et al., 2020).

Para problemas con clases no equilibradas en conjuntos de datos, existe una aplicación del Coeficiente de Correlación de Matthews. Es la única métrica que tiene en cuenta todos los elementos de una matriz de confusión, por lo que no importa el desequilibrio extremo entre clases, da valores equilibrados y fiables. Según (Chollet, 2021), MCC es una métrica que refleja con mayor precisión la calidad de las predicciones cuando las clases no todas tienen la misma cantidad de datos.

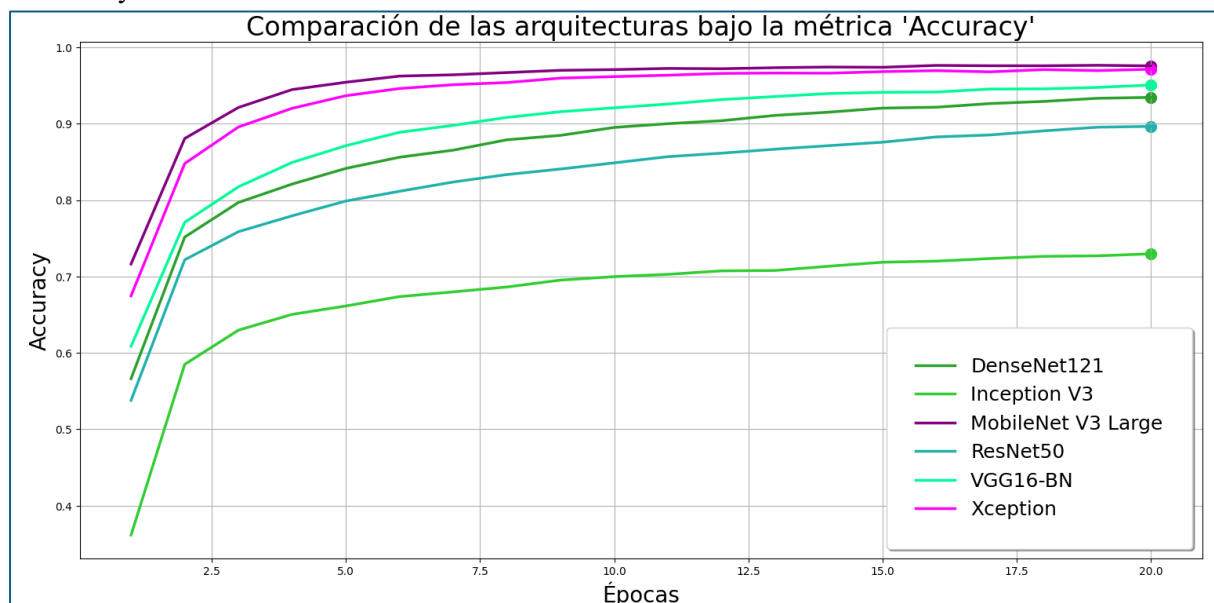
RESULTADOS Y DISCUSIÓN

En la Figura 1 se presenta el rendimiento de diferentes arquitecturas de redes neuronales convolucionales (CNN), utilizando la métrica de precisión (Accuracy). En términos del rendimiento general se observa que las arquitecturas Xception y MobileNet V3 alcanzan la mayor precisión en comparación con los otros modelos. DenseNet121 muestra un desempeño inferior alcanzando el valor

más bajo de precisión. Los modelos Inception V3, ResNet50 y VGG16 tienen curvas de aprendizaje constantes, aunque no tan rápidas ni precisas comparadas con los modelos Xception y MobileNet V3. La mayoría de las arquitecturas logran estabilizar su aprendizaje entre 10 y 15 épocas.

En base a la métrica “Accuracy” las opciones Xception y MobileNet V3 son las más adecuadas, haciendo un balance entre precisión y eficiencia, alcanzando niveles altos de rendimiento con menor número de épocas.

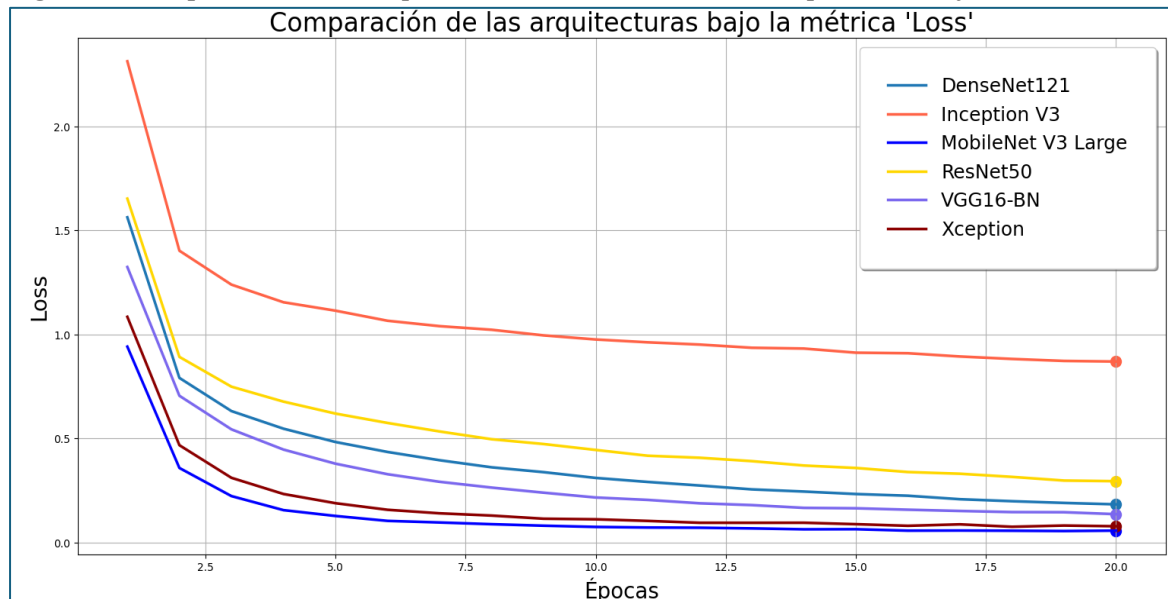
Figura 1: Comparación de las arquitecturas, sin modificación de arquitectura, bajo la métrica “Accuracy”



El análisis de pérdida es fundamental ya que indica que tan bien el modelo está aprendiendo a partir de los datos al minimizar los errores de predicción. En la Figura 2 se muestra que en todas las arquitecturas la pérdida disminuye a medida que avanzan las épocas. La tasa de convergencia y los valores finales de pérdida varían significativamente entre las arquitecturas. Xception y MobileNet V3 presentan las pérdidas más bajas al final del entrenamiento, las curvas son suaves y hay una rápida reducción de pérdida en las primeras épocas.

Por otra parte, Inception V3 muestra una curva de pérdida más lenta en comparación con los otros modelos, su valor de pérdida es considerablemente más alto, esto podría indicar limitaciones en su capacidad para clasificar correctamente con este conjunto de datos. Los modelos DenseNet121, ResNet50 y VGG16 presentan curvas intermedias, alcanzando valores aceptables, pero sin igualar al desempeño de Xception y MobileNet V3.

Figura 2: Comparación de las arquitecturas, sin modificación de arquitectura, bajo la métrica “Loss”



La Tabla 4 muestra los resultados de las arquitecturas bajo la mejor configuración. En este caso, la arquitectura más eficaz de entre todas las configuraciones es MobileNetV3, especialmente con un tamaño de lote de 32 y una tasa de aprendizaje de 1×10^{-4} , el modelo alcanzó una precisión de prueba del 96,89%, registró la pérdida más baja con un 11,38%, mantuvo las métricas “*precisión*”, “*recall*”, “*F1 score*” y MCC por encima de 95%. Este rendimiento continuo demuestra que el modelo se adapta bien a las condiciones cambiantes y, por tanto, mejora su capacidad de generalización al tiempo que mantiene su eficiencia operativa.

Tabla 4: Evaluación de sistema de métricas para las arquitecturas.

Model	Training Time	Validation Accuracy	Test Accuracy	Test Loss	Test Precision	Test Recall	Test F1 Score	Test MCC
Xception	3568s	96.41%	94.99%	22.51%	94.17%	94.17%	94.14%	93.97%
MobileNet	3520s	96.89%	96.49%	11.38%	95.37%	95.33%	95.33%	95.17%
InceptionV3	4493s	86.52%	82.80%	43.94%	83.35%	82.80%	82.44%	82.25%
DenseNet121	5072s	96.24%	95.06%	17.84%	94.01%	93.90%	93.87%	93.70%
ResNet50	3709s	94.07%	91.80%	25.16%	92.24%	91.80%	91.67%	91.54%
VGG16-BN	3433s	96.54%	95.27%	14.81%	95.31%	95.27%	95.23%	95.11%

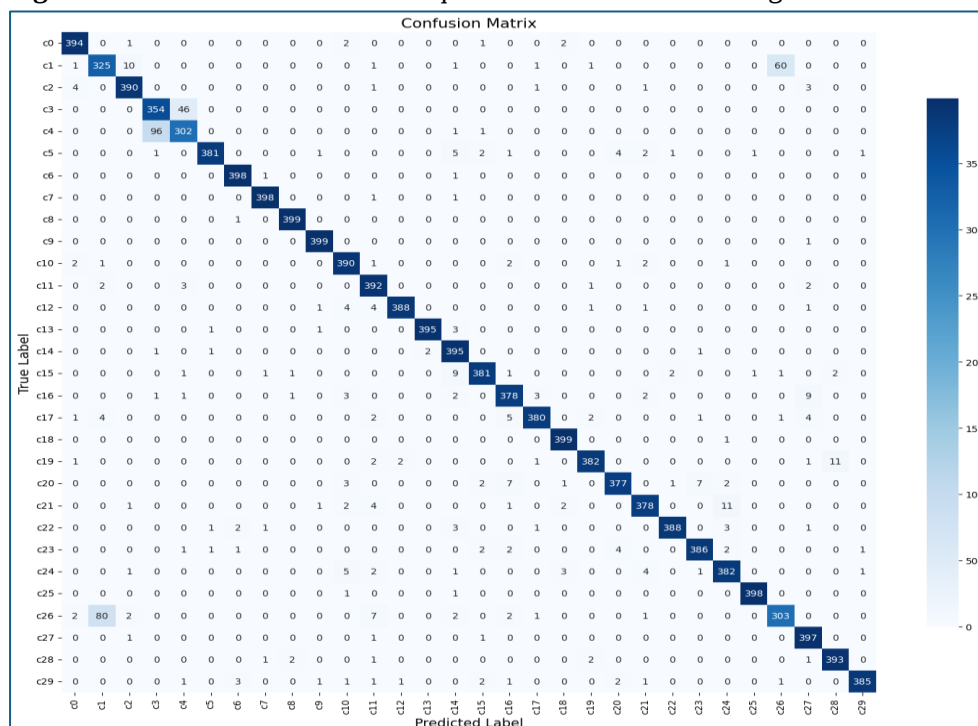
Por otro lado, algunas arquitecturas, como DenseNet121 y Xception, presentan características de rendimiento competitivas, pero algo inestables. Por ejemplo, DenseNet121 alcanza una precisión de prueba del 96,24% con un bajo índice de pérdidas del 17,84%; después sufre caídas en la métrica de

validación bajo diferentes configuraciones. Los modelos InceptionV3, ResNet50 y VGG16-BN tienen métricas menos favorables, que se caracterizan por mayores pérdidas y menor precisión de prueba, mostrando así limitaciones en su capacidad de generalización a diferentes configuraciones. Esto sirve para demostrar que MobileNetV3 no sólo destaca en métricas específicas, sino que también consigue mantener un rendimiento constante en toda una gama de configuraciones, por lo que es más resistente.

Arquitectura destacada

La matriz de confusión de MobileNetV3 Large FT (Ver Figura 3), muestra un buen rendimiento para la mayoría de las clases, con valores altos a lo largo de la diagonal, lo que indica que el modelo realiza predicciones precisas. Las clasificaciones erróneas son mínimas; sin embargo, clases como Latas de comida de acero [c26] revelan algunas confusiones notables, principalmente con Latas de comida de aluminio [c1], con 80 casos clasificados erróneamente. Del mismo modo, las bolsas de la compra de plástico [c20] y las bolsas de basura de plástico [c23] comparten algunas clases -probablemente debido a las mismas características visuales- y, por lo tanto, clasifican erróneamente algunas muestras. No obstante, MobileNetV3 Large FT demuestra realmente una gran precisión de clasificación, demostrando que puede diferenciar la mayoría de las clases de residuos con sólo una confusión menor en las clases que comparten características visuales similares.

Figura 3: Matriz de confusión de arquitectura MobileNetV3 Large.



Sistema de métricas por clase de Arquitectura destacada

Tabla 5: Sistema de métricas por clase de arquitectura MobileNet V3 Large

Class	Accuracy	Precision	Recall	F-Score	MCC
Latas de aerosol	1.00	0.99	0.99	0.99	0.95
Latas de aluminio para alimentos	0.99	0.90	0.73	0.81	0.77
Latas de refresco de aluminio	0.99	0.97	0.98	0.97	0.92
Cajas de cartón	0.99	0.86	0.88	0.87	0.78
Envases de cartón	0.99	0.86	0.85	0.86	0.78
Ropa	0.99	1.00	0.97	0.98	0.90
Posos de café	1.00	0.99	1.00	0.99	0.97
Cubiertos de plástico desechables	1.00	0.99	1.00	0.99	0.96
Cáscaras de huevo	1.00	0.98	0.98	0.98	0.97
Residuos de alimentos	1.00	1.00	0.97	0.98	0.98
Botellas de vidrio para bebidas	0.99	0.98	0.99	0.98	0.93
Envases de vidrio para bebidas	1.00	0.99	0.98	0.98	0.95
Tarros de vidrio para alimentos	1.00	0.99	0.99	0.99	0.97
Revistas	1.00	0.98	0.99	0.99	0.93
Periódicos	0.99	0.97	0.99	0.98	0.90
Papel de oficina	0.99	0.99	1.00	0.99	0.92
Vasos de papel	0.99	0.98	0.95	0.97	0.89
Tapas de plástico	0.99	0.98	0.98	0.98	0.90
Botellas de plástico de detergente	1.00	1.00	1.00	1.00	0.98
Envases de plástico para alimentos	1.00	0.99	0.98	0.98	0.94
Bolsas de plástico	1.00	1.00	0.97	0.98	0.95
Botellas de plástico de refresco	0.99	0.99	0.96	0.98	0.91
Pajitas de plástico	1.00	0.96	0.99	0.98	0.96
Bolsas de basura de plástico	1.00	0.99	0.97	0.98	0.94
Botellas de agua de plástico	0.99	0.92	0.98	0.95	0.90
Zapatos	1.00	0.99	0.98	0.99	0.98
Latas de acero	0.99	0.79	0.93	0.86	0.82
Vasos de poliestireno	1.00	0.97	1.00	0.98	0.97
Contenedores de comida	1.00	0.98	0.97	0.98	0.95
Bolsas de té	0.99	0.99	0.98	0.99	0.89
Promedio	0.99	0.96	0.96	0.96	0.91

La tabla 5 ilustra el rendimiento de MobileNetV3 Large FT en cada clase de residuos, con unos valores medios de exactitud, precisión, recuperación y puntuación F del 99%, 96%, 96% y 96%, respectivamente, con un MCC del 91%, lo que indica una clasificación general sólida. Entre las mejores clases en términos de puntuación perfecta en todas las métricas se encuentran las botellas de detergente de plástico, los zapatos y los cubiertos de plástico desechables, para estas clases se obtienen muy buenos resultados en términos de precisión del modelo. Sin embargo, las latas de comida de acero y las latas de comida de aluminio tienen valores más bajos de precisión (79% y 90%, respectivamente) y MCC, lo que refleja la dificultad para distinguirlas de clases similares. En general, MobileNetV3 Large FT demuestra una gran precisión y un rendimiento equilibrado.

CONCLUSIONES

MobileNetV3 Large FT es la mejor arquitectura para tareas de clasificación de residuos, superando a otros modelos CNN en precisión, eficiencia y adaptabilidad. Tiene la mejor precisión de prueba con un 96,49% y la menor pérdida con un 11,38%, lo que demuestra un mejor rendimiento tanto en precisión de clasificación como en minimización de errores. También los valores de “*precisión*”, “*recall*”, “*score*” y MCC son mayores del 95%, lo que demuestra un rendimiento estable y fiable en una amplia gama de tipos de residuos.

La matriz de confusión de MobileNetV3 Large FT muestra que la mayoría de las clases de residuos se clasifican con gran precisión, y que sólo se producen ligeros errores de clasificación entre clases con características visuales similares, como latas de comida de acero y latas de comida de aluminio. Esto se ve reforzado por la Tabla 5, donde en muchas categorías, como las botellas de detergente de plástico, los zapatos y los cubiertos de plástico desechables, los resultados son casi perfectos.

La eficiencia de MobileNetV3 Large FT es alta, a pesar del tamaño reducido del modelo, además, el diseño arquitectónico está desarrollado para ser ejecutados en dispositivos sin dependencia a CUDA. Esto la hace muy adecuada para aplicaciones prácticas en las que la solución requerida debe ser rápida, precisa y eficiente en el uso de los recursos para ajustarse a los objetivos de una gestión sostenible de los residuos. En resumen, MobileNetV3 Large FT muestra el mejor compromiso entre eficiencia computacional y precisión en la clasificación, por tanto, constituye la elección óptima para la clasificación de residuos.

REFERENCIAS BIBLIOGRAFICAS

- Atienza, Rowel. (2020). *Advanced Deep Learning with TensorFlow 2 and Keras* (J. Gonsalves, Dr. I. Hough, & K. Sonawane, Eds.; Second Edition). Packt Publishing. <https://www.packtpub.com/en-us/product/advanced-deep-learning-with-tensorflow-2-and-keras-9781838821654>
- Ayyadevara, K., & Reddy, Y. (2020). *Modern Computer Vision with PyTorch: Explore deep learning concepts and implement over 50 real-world image applications* (S. Shetty, S. Mandal, J. Sunil, R. Kumar, M. Kurup, & S. Editing, Eds.; First edition). Packt Publishing Ltd. <https://www.packtpub.com/en-us/product/modern-computer-vision-with-pytorch-9781839213472>
- Bianco, S., Cadene, R., Celona, L., & Napoletano, P. (2018). Benchmark Analysis of Representative Deep Neural Network Architectures. *IEEE Access*, 4, 8. <https://doi.org/10.1109/ACCESS.2017.DOI>
- Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., & Lindauer, M. (2021). *Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges*. <https://doi.org/10.1002/widm.1484>
- Chollet, F. (2021). *Deep Learning with Python* (J. Stout, Ed.; Second). Manning Publications.
- Da, E., Puls, S., Todescato, M. V., & Carbonera, J. L. (2023). AN EVALUATION OF PRE-TRAINED MODELS FOR FEATURE EXTRACTION IN IMAGE CLASSIFICATION. 32. <https://doi.org/10.48550>
- Dawani, Jay. (2020). *Hands-on mathematics for deep learning : build a solid mathematical foundation for training efficient deep neural networks* (S. Shetty, A. Abidi, A. Sapuni, R. Kumar, M. Kurup, & S. Editing, Eds.). Packt Publishing. <https://www.packtpub.com/en-us/product/hands-on-mathematics-for-deep-learning-9781838647292>
- Elgendy, Mohamed. (2020). *Deep Learning for Vision Systems* (J. Stout, A. Courniot, I. Martinović, L. Weidert, & T. Taylor, Eds.; 1st edition). Manning Publications. <https://www.manning.com/books/deep-learning-for-vision-systems>
- Ilievski, I., Akhtar, T., Feng, J., & Shoemaker, C. A. (2017). Efficient Hyperparameter Optimization of



- Deep Learning Algorithms Using Deterministic RBF Surrogates. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31, 8. <https://doi.org/10.1609/aaai.v31i1.10647>
- Koul, A., Ganju, S., & Kasam, M. (2019). *Practical Deep Learning for Cloud, Mobile, & Edge Real-World AI and Computer-Vision Projects Using Python, Keras, and TensorFlow* (R. Roumeliotis, N. Tache, & C. Faucher, Eds.; First Edition). O'Reilly Media. <http://oreilly.com>
- Kunwar, S. (2023). *MWaste: A Deep Learning Approach to Manage Household Waste*. <https://doi.org/10.20944/preprints202304.0066.v1>
- Morales M. (2020). *Grokking Deep Reinforcement Learning* (J. Stout, A. Krinker, I. Martinović, & K. Petito, Eds.; First Edition). Manning Publications Co.
- Narayan, Y. (2021, January 15). DeepWaste: Applying Deep Learning to Waste Classification for a Sustainable Planet. *V 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada. <https://doi.org/10.48550/arXiv.2101.05960>
- Novello, P., Poëtte, G., Lugato, D., & Congedo, P. M. (2023). Goal-Oriented Sensitivity Analysis of Hyperparameters in Deep Learning. *Journal of Scientific Computing*, 94. <https://doi.org/10.1007/s10915-022-02083-4>
- Stevens, E., Antiga, L., & Viehmann, T. (2020). *Deep Learning with PyTorch* (Simon and Schuster, Ed.; First Edition). Manning.
- White, G., Cabrera, C., Palade, A., Li, F., & Clarke, S. (2020). *WasteNet: Waste Classification at the Edge for Smart Bins*. 8. <https://doi.org/10.48550/arXiv.2006.05873>