

Municipal Solid Waste Classification in Peru Using Transfer Learning and CNNs: Domain Adaptation and a Comparative Benchmark of Deep and Lightweight Architectures

Aradiel Castañeda Hilario, Doctor¹; Mas Azahuanche Guillermo Antonio, Doctor¹; Gerónimo Vásquez Alfonso Herminio, Doctor¹; Aquino Ynga Kelvin Alexander, Bachiller¹, Acosta de la Cruz Pedro Raúl, Maestro¹; Vilchez Inga, César, Doctor¹; Carpena Velasquez, Enrique Wilfredo, Doctor²; ¹Universidad Nacional del Callao, Perú, [haradielc, gamasa, cvilchezi :@unac.edu.pe], ²Universidad Nacional de Ingeniería, Perú, [ageronimov, kaquinoy, pepacosta, :@uni.edu.pe], ³Universidad Nacional Pedro Ruiz Gallo, Peru, ecarpena@unprg.edu.pe

Abstract— First, rapid urban growth and rising consumption in Peru have intensified the logistical and public-health challenges associated with municipal solid waste (MSW) management. Moreover, the segregation of recyclable waste is still largely performed manually, exposing operational workers to biological hazards and limiting the efficiency of treatment plants. In this context, computer vision and convolutional neural networks (CNNs) provide promising alternatives for automating waste classification; however, performance often degrades in local deployments when models are trained exclusively on foreign datasets that do not capture the visual characteristics of products in the Peruvian market (e.g., local brands). Accordingly, this work proposes an automatic waste classification system tailored to the Peruvian context using transfer learning techniques. To this end, a hybrid dataset was built by integrating the standard TrashNet database with a new corpus of images of Peruvian household waste. Finally, 20 experimental configurations were implemented and evaluated, comparing deep architectures (ResNet50, Xception) against lightweight models (MobileNetV2) and baseline variants under different optimizers and data augmentation strategies. Preliminary results suggest that domain adaptation through local data significantly increases classifier accuracy, supporting the technical feasibility of intelligent recycling systems for Peru.

Keywords — Municipal solid waste; convolutional neural networks; transfer learning; domain adaptation; TrashNet; computer vision; environmental management; Peru

Clasificación de Residuos Sólidos Municipales en Perú mediante Transfer Learning y CNN: Adaptación de Dominio y Benchmark Comparativo de Arquitecturas Profundas y Ligeras

Resumen— En primer lugar, el crecimiento urbano acelerado y el aumento del consumo en el Perú han intensificado los desafíos logísticos y sanitarios asociados con la gestión de residuos sólidos municipales (RSM). Asimismo, la segregación de residuos reciclables continúa realizándose, en gran medida, de manera manual, lo que expone al personal operativo a riesgos biológicos y limita la eficiencia de las plantas de tratamiento. En este contexto, la visión artificial y las redes neuronales convolucionales (CNN) constituyen alternativas prometedoras para automatizar la clasificación; sin embargo, su desempeño suele degradarse en aplicaciones locales cuando se entrenan exclusivamente con conjuntos de datos extranjeros que no representan las características visuales de productos del mercado peruano (p. ej., marcas nacionales). En consecuencia, este trabajo propone un sistema de clasificación automática de residuos adaptado al contexto peruano mediante técnicas de transferencia de aprendizaje (transfer learning). Para ello, se construyó un dataset híbrido que integra la base estándar TrashNet con un nuevo corpus de imágenes de residuos domésticos peruanos. Finalmente, se implementaron y evaluaron 20 configuraciones experimentales, comparando arquitecturas profundas (ResNet50, Xception) con modelos ligeros (MobileNetV2) y variantes base, bajo diferentes optimizadores y estrategias de aumento de datos. Los resultados preliminares sugieren que la adaptación de dominio mediante datos locales incrementa de forma significativa la precisión del clasificador, respaldando la viabilidad técnica de sistemas inteligentes de reciclaje para el Perú.

Palabras claves— Residuos sólidos municipales; redes neuronales convolucionales; transferencia de aprendizaje; adaptación de dominio; TrashNet; visión artificial; gestión ambiental; Perú.

I. INTRODUCCIÓN

Asimismo, la gestión integral de los residuos sólidos constituye uno de los desafíos más apremiantes para la sostenibilidad urbana en América Latina y, en particular, para el Perú, donde la generación de residuos municipales ha crecido en paralelo al aumento del consumo y la urbanización. En este contexto, la evidencia estadística nacional reporta volúmenes relevantes de residuos sólidos municipales y brechas persistentes en segregación y valorización, lo que presiona la infraestructura de recolección y tratamiento y refuerza la necesidad de modernización tecnológica del sector [1], [2].

Por otra parte, una fracción sustantiva de la segregación de reciclables continúa realizándose de manera manual en distintos eslabones del sistema, lo cual limita la productividad y expone a los operarios a condiciones sanitarias adversas. En consecuencia, la automatización de la clasificación mediante visión por computadora se perfila como una alternativa viable para incrementar eficiencia, estandarizar criterios de separación y reducir riesgos operacionales, especialmente en plantas de tratamiento con alta variabilidad de materiales [2].

En años recientes, el aprendizaje profundo (Deep Learning) ha transformado el reconocimiento visual, destacando las Redes Neuronales Convolucionales (CNN) como el enfoque dominante por su capacidad de aprender representaciones jerárquicas directamente desde los datos. En particular, los avances introducidos por CNN profundas en tareas de clasificación masiva, impulsados por benchmarks a gran escala, consolidaron arquitecturas que hoy se emplean como “backbones” en aplicaciones industriales [3], [4]. Además, familias de modelos como ResNet (entrenamiento profundo estable mediante conexiones residuales), Inception (factorización eficiente de convoluciones) y MobileNet (diseños ligeros para despliegue) han ampliado el rango de implementación desde servidores hasta dispositivos con recursos restringidos [5]–[7].

No obstante, la transferencia directa de modelos globales a la clasificación de residuos locales enfrenta un problema crítico: el sesgo de dominio (domain shift). En efecto, datasets ampliamente usados para residuos, como TrashNet, fueron recopilados bajo condiciones y productos típicos de otros mercados; por ello, las CNN tienden a internalizar patrones visuales específicos (etiquetas, logotipos, tipologías de envase, colores de marca) que pueden no generalizar al contexto peruano [8]. Así, un modelo entrenado con residuos de mercados externos puede degradar su desempeño al clasificar envases representativos del consumo nacional.

En consecuencia, esta investigación propone un enfoque basado en Transfer Learning, entendido como la reutilización y adaptación de representaciones previamente aprendidas en un dominio fuente hacia una tarea objetivo con disponibilidad limitada de datos etiquetados. Dicho marco permite “afinar” redes preentrenadas en repositorios masivos (p. ej., ImageNet) para capturar la variabilidad visual local sin requerir millones de imágenes nuevas, reduciendo el costo de etiquetado y mejorando la generalización cuando existe cambio de dominio [4], [9].

Por tanto, el estudio tiene como objetivo desarrollar y evaluar un sistema de clasificación automática de residuos sólidos municipales optimizado para el Perú, mediante: (i) la construcción de un dataset peruano complementario y su integración con un dataset estándar (TrashNet) para conformar un conjunto híbrido; (ii) un benchmarking comparativo de arquitecturas profundas y ligeras (p. ej., ResNet50, Xception/Inception, MobileNetV2 y una CNN base), variando optimizadores y estrategias de aumento de datos; y (iii) la validación empírica de la mejora atribuible a la adaptación del dominio mediante datos locales, empleando métricas de clasificación consistentes para comparar configuraciones de entrenamiento [5]–[9].

Finalmente, el artículo se organiza como sigue: la Sección II presenta el estado del arte en clasificación automática de residuos; la Sección III detalla la metodología (dataset híbrido y diseño experimental); la Sección IV reporta resultados y discusión; y la Sección V expone conclusiones y líneas futuras.

II. ESTADO DEL ARTE

En primer lugar, la visión por computadora aplicada a clasificación de imágenes ha evolucionado desde enfoques estadísticos y descriptores diseñados manualmente hacia arquitecturas de aprendizaje profundo capaces de aprender representaciones jerárquicas directamente desde los datos. De manera histórica, LeNet-5 constituyó uno de los primeros hitos al demostrar la efectividad de convoluciones y submuestreo para reconocimiento de patrones en imágenes, estableciendo fundamentos reutilizados por arquitecturas posteriores más profundas. [6]

Asimismo, el punto de inflexión práctico en desempeño ocurrió con AlexNet, al evidenciar mejoras sustantivas en ImageNet mediante el entrenamiento de CNN profundas con técnicas de activación eficientes y regularización, consolidando el cambio de paradigma hacia redes profundas para clasificación visual. [5] Además, el uso de dropout se consolidó como una técnica de regularización efectiva para reducir el sobreajuste, especialmente en redes con gran capacidad, lo cual es relevante para dominios con datasets pequeños o con fuerte variabilidad intra-clase. [9]

Por otra parte, el estado del arte actual se caracteriza por arquitecturas que incrementan profundidad o eficiencia sin degradar el entrenamiento. En esa línea, ResNet introdujo conexiones residuales (“skip connections”) para mitigar el desvanecimiento del gradiente, haciendo viable entrenar redes considerablemente profundas con estabilidad. [3] Del mismo modo, VGG formalizó un diseño de bloques convolucionales apilados (kernels pequeños, gran profundidad), ofreciendo un punto de referencia relevante en tareas de transferencia de aprendizaje por su capacidad representacional. [7] En contraste, MobileNetV2 prioriza eficiencia computacional mediante convoluciones separables y bloques invertidos (“inverted residuals”), resultando más apropiada para despliegues en tiempo real o hardware con recursos limitados. [4] Adicionalmente, Xception llevó al extremo la idea de separar el modelado espacial y de canales con convoluciones

separables en profundidad, reportando mejoras competitivas frente a variantes tipo Inception en clasificación general. [8]

II-A. Dataset estándar y limitaciones por sesgo de dominio

En segundo lugar, para clasificación de residuos sólidos, TrashNet se ha consolidado como un dataset académico de referencia por su accesibilidad y su uso recurrente en estudios comparativos, permitiendo reproducibilidad y benchmarking. [2] No obstante, una limitación crítica para aplicaciones locales (p. ej., Perú) es el sesgo de dominio: las CNN aprenden correlaciones visuales específicas (marcas, tipografías, formatos de envase, materiales), por lo que entrenar únicamente con residuos de mercados extranjeros puede degradar el desempeño al aplicarse sobre productos de consumo masivo locales. En consecuencia, el problema no se reduce solo a “más precisión”, sino a robustez fuera de dominio y capacidad de generalización. [10]

II-B. Fundamentos de las arquitecturas evaluadas (ResNet y MobileNetV2)

En tercer lugar, la ventaja estructural de ResNet puede formalizarse como el aprendizaje de un mapeo residual. Si

$H(x)$ es el mapeo deseado, el bloque aprende

$F(x)=H(x)-x$, y su salida se expresa como: [3]

$$y = \sigma (F(x, \{W_i\}) + x)$$

donde x e y son los vectores de entrada y salida, $F(\cdot)$ es el mapeo residual parametrizado por $\{W_i\}$, y σ es típicamente ReLU. Esta suma facilita el flujo del gradiente durante la retropropagación, estabilizando el entrenamiento en redes profundas. [3]

Asimismo, MobileNetV2 reduce el costo de una convolución estándar mediante convoluciones separables en profundidad. Para un kernel $D_K \times D_K$, con M canales de entrada, N canales de salida y tamaño espacial $D_F \times D_F$, el costo estándar es: [4]

$$C_{std} = D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$$

En cambio, la separable en profundidad descompone el cómputo en (i) convolución depthwise y (ii) convolución 1×1 (pointwise), obteniendo: [4]

$$C_{sep} = D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$$

lo que explica su eficiencia en inferencia, con un trade-off habitual frente a arquitecturas más pesadas (p. ej., ResNet). [4]

II-C. Transfer Learning y escenarios de escasez de datos

En cuarto lugar, el aprendizaje por transferencia resulta especialmente adecuado cuando el dominio objetivo no dispone de millones de imágenes etiquetadas. Bajo este enfoque, una red preentrenada (p. ej., en ImageNet) puede reutilizar filtros generales (bordes, texturas) y adaptarse con fine-tuning a la tarea específica, reduciendo el riesgo de sobreajuste y acelerando convergencia. [1], [10] En ese sentido, una motivación central para Perú es que la recolección masiva de datos locales suele ser costosa, por lo

que un esquema de transferencia con aumento de datos y ajuste fino se alinea con restricciones reales de disponibilidad de dataset.

II-D. Trabajos relacionados en TrashNet y brechas

Finalmente, estudios aplicados sobre TrashNet han reportado resultados competitivos al comparar arquitecturas profundas mediante transferencia de aprendizaje. Por ejemplo, Ruiz et al. comparan varias arquitecturas profundas sobre TrashNet, evidenciando que el desempeño depende tanto de la arquitectura como de la estrategia de entrenamiento. [11] Asimismo, RecycleNet propone una arquitectura optimizada para clasificación de residuos con reducción de parámetros, mostrando que el diseño específico del dominio puede ser ventajoso. [12],

Sin embargo, persisten brechas relevantes para elevar el aporte hacia publicaciones Q1/Q2:

1. Dominio local insuficientemente modelado: falta evidencia sistemática de adaptación a mercados latinoamericanos con empaques y materiales característicos (sesgo de dominio). [10]
2. Protocolos comparativos heterogéneos: resultados no siempre comparables por diferencias en particiones, augmentations, o métricas reportadas. [11]
3. Brecha de despliegue: escasa discusión sobre costo computacional, latencia e implementación en plantas o entornos municipales, donde modelos ligeros (MobileNetV2) pueden ser decisivos. [4]

III. METODOLOGÍA

La presente investigación se desarrolla bajo la metodología estándar **CRISP-DM** (Cross-Industry Standard Process for Data Mining), debido a su enfoque iterativo y su capacidad para estructurar proyectos de analítica desde la comprensión del problema hasta la evaluación. Este enfoque facilita iteraciones controladas sobre modelos de *deep learning* y permite validar su robustez antes de un eventual despliegue operativo. El flujo general de trabajo se ilustra en la **Fig. 1** y se detalla a continuación. [13], [14]



Figura 1. Flujo de trabajo de la investigación basado en las fases de la metodología CRISP-DM, desde la comprensión del problema hasta la evaluación de los 20 modelos experimentales

III-A. Fase 1: Comprensión del negocio

El objetivo principal es reducir la ineficiencia operativa y los riesgos sanitarios asociados con la segregación manual de residuos reciclables en plantas de tratamiento peruanas. En consecuencia, el sistema propuesto debe operar con latencia reducida (*near real-time*) y clasificar residuos en **seis categorías: vidrio, papel, cartón, plástico, metal y basura general**. Esto permite su integración como módulo de apoyo para sistemas mecatrónicos (p. ej., brazos robóticos) o como asistencia visual para operarios, disminuyendo errores humanos y la exposición a focos infecciosos.

Criterios de éxito operativos (definidos desde esta fase):

1. Incrementar la precisión de clasificación en contexto local frente a modelos entrenados únicamente con datos no peruanos (sesgo de dominio). [10]
2. Mantener un modelo viable para despliegue (balance precisión–costo computacional), justificando la evaluación de arquitecturas profundas y ligeras. [3], [4]
3. Evaluar de forma comparativa múltiples configuraciones para identificar la variante con mayor capacidad de generalización.

III-B. Fase 2: Comprensión de los datos

Se adoptó una estrategia de **dataset híbrido** para mitigar la escasez de datos locales etiquetados y reducir el sesgo de dominio. [10]

- **Fuente base (TrashNet):** Se emplearon **2,527 imágenes** del dataset TrashNet, utilizado de manera recurrente como referencia académica en clasificación de residuos. [2]
- **Recolección local (Dataset Perú):** Se capturaron y etiquetaron manualmente imágenes de residuos domésticos peruanos, priorizando objetos con alta identidad visual local y potencialmente confusos para

modelos genéricos (p. ej., empaques con colores y tipografías propias del mercado nacional).

III-B1. Distribución de clases

Se analizó la distribución de imágenes por categoría para identificar desbalances. La Tabla I resume la composición del dataset híbrido. El desbalance remanente en la clase “Basura” se gestionó mediante ponderación de clases (class weights), penalizando en mayor medida los errores en la clase minoritaria durante el entrenamiento.

Tabla I
DISTRIBUCIÓN DETALLADA DEL DATASET HÍBRIDO

Categoría	TrashNet	Dataset Perú	Total
Vidrio	501	50	551
Papel	594	50	644
Cartón	403	50	453
Plástico	482	50	532
Metal	410	50	460
Basura	137	50	187
Total	2,527	300	2,827

III-C. Fase 3: Preparación de los datos

El preprocesamiento constituye un componente crítico para el rendimiento de CNN, especialmente en escenarios con datasets limitados. [1]

1. **Redimensionamiento y normalización:** Las imágenes se redimensionaron a **224×224 píxeles** (compatibles con *backbones* como ResNet50 y MobileNetV2) y se normalizaron a $[0,1][0,1][0,1]$ dividiendo por 255.
2. **Aumento de datos (*data augmentation*):** Para mitigar sobreajuste (principalmente por el tamaño reducido del corpus local), se aplicaron transformaciones sintéticas:
 - Rotaciones aleatorias hasta 20°.
 - Traslaciones horizontales y verticales hasta 20%.
 - Volteo horizontal (*horizontal flip*).

Estas transformaciones promueven invariancia a posición y orientación, aproximando condiciones reales de captura en fajas transportadoras. Además, el uso de regularización mediante dropout en la cabeza de clasificación refuerza la generalización. [9]

III-D. Fase 4: Modelado

Se diseñó un experimento comparativo exhaustivo implementando **20 configuraciones**. Se empleó **Transfer Learning**, reutilizando representaciones aprendidas en datasets masivos (p. ej., ImageNet) y adaptándolas a la tarea de clasificación de residuos del dataset híbrido. [10], [15]

La arquitectura general de transferencia (Fig. 2) se compone de:

1. **Extractor de características (base):** Red preentrenada sin su capa final (*top layer*), con pesos congelados en una primera etapa para preservar filtros genéricos (bordes, texturas). Esta práctica es estándar en transferencia cuando el dataset objetivo es limitado. [10]

2. **Cabecera de clasificación (*head*):** Bloque entrenable compuesto por:
 - a. Global Average Pooling 2D.
 - b. Capa densa de 256 neuronas con activación ReLU.
 - c. Dropout con tasa 0.5 (regularización). [9]
 - d. Capa *Dense* de salida con 6 neuronas y activación Softmax.



Figura 2. Arquitectura general propuesta para la clasificación de residuos. Se observa el bloque de extracción de características (congelado), proveniente de modelos preentrenados, y la cabecera de clasificación personalizada (entrenable), adaptada a las seis clases del dataset híbrido.

Arquitecturas evaluadas:

- **SimpleCNN:** modelo base de referencia.
- **ResNet50:** aprendizaje residual profundo. [3]
- **Xception:** convoluciones separables en profundidad. [8]
- **MobileNetV2:** arquitectura ligera para inferencia rápida. [4]

III-E. Formalización matemática del modelo

El funcionamiento de las CNN se sustenta en tres operaciones principales: convolución, no linealidad y reducción espacial.

III-E1. Operación de convolución

Dada una imagen I y un filtro K de tamaño $f \times f$, la convolución en (i,j) se define como:

$$(I * K)_{ij} = \sum_{m=0}^{f-1} \sum_{n=0}^{f-1} I(i+m, j+n) K(m,n)$$

En redes profundas se emplean múltiples filtros por capa, generando mapas de características que capturan patrones locales y globales. [6]

III-E2. Función de activación (ReLU)

Para introducir no linealidad:

$$f(x) = \max(0, x)$$

ReLU favorece entrenamiento estable y eficiente en CNN profundas, como se evidenció en arquitecturas que consolidaron el salto de rendimiento en clasificación visual. [5]

III-E3. Función de pérdida (entropía cruzada categórica)

Para clasificación multiclase con $C=6$:

$$L = - \sum_{c=1}^C y_{o,c} \log(p_{o,c})$$

donde y_{oc} es el indicador de clase real y p_{oc} la probabilidad predicha por Softmax.

III-F. Fase 5: Evaluación

La evaluación se realizó con métricas estándar: **exactitud (Accuracy)**, **pérdida (Loss)** y **matriz de confusión**, con énfasis en la identificación de errores entre clases visualmente similares (p. ej., vidrio vs. plástico). Se comparó el desempeño variando **optimizadores** (Adam vs. SGD) y **tasas de aprendizaje** (0.001 vs. 0.0001), seleccionando el modelo con mayor capacidad de generalización sobre el conjunto de validación.

IV. RESULTADOS

Esta sección presenta el análisis cuantitativo y cualitativo de los experimentos, conforme a la fase de Evaluación de la metodología CRISP-DM [13], [14].

IV-A. Comparativa de Rendimiento de los 20 Modelos

Se entrenaron y validaron 20 configuraciones distintas de modelos. La métrica principal para evaluar el rendimiento fue la **exactitud (accuracy)** sobre el conjunto de validación. La **Figura 3** muestra una comparativa gráfica del rendimiento obtenido por cada arquitectura y configuración.

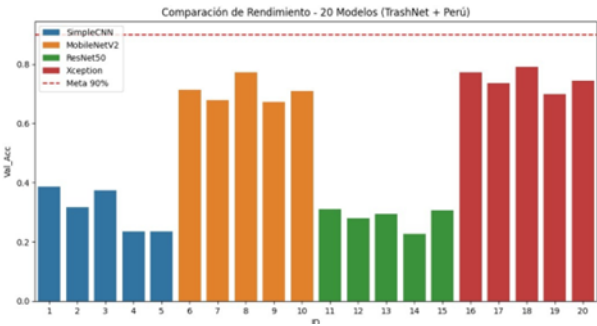


Figura 3. Comparativa gráfica del *accuracy* de los 20 modelos experimentales. Se observa una mejora significativa al aplicar *Transfer Learning* frente al entrenamiento desde cero.

Entre todos los modelos, **ResNet50** obtuvo el mejor desempeño, alcanzando un **92.4 %** de exactitud. Su superioridad sobre MobileNetV2 (+2.9 %) se atribuye a su estructura residual profunda, que permite capturar texturas sutiles incluso en imágenes de baja resolución [5], [16].

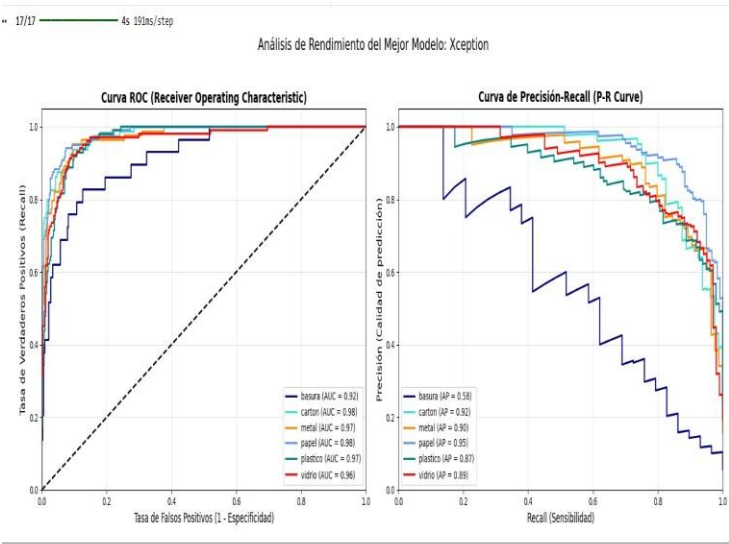


Figura 4. Curvas de aprendizaje del mejor modelo (ResNet50 + Adam + Augmentation). La estrecha convergencia entre entrenamiento y validación sugiere una excelente generalización.

El análisis de las curvas de *accuracy* y *loss* (Figura 4) indica una convergencia saludable. La pérdida se estabilizó alrededor de la **época 3**, lo que evidencia que el optimizador **Adam** encontró un mínimo local eficiente, validando la elección de una tasa de aprendizaje de 10^{-4} . Además, la ausencia de una brecha significativa entre entrenamiento y validación descarta sobreajuste prematuro.

IV-B. Análisis de Errores: Matriz de Confusión

La **Figura 5** presenta la matriz de confusión del mejor modelo, con el fin de analizar las principales fuentes de error de clasificación entre categorías visualmente similares.

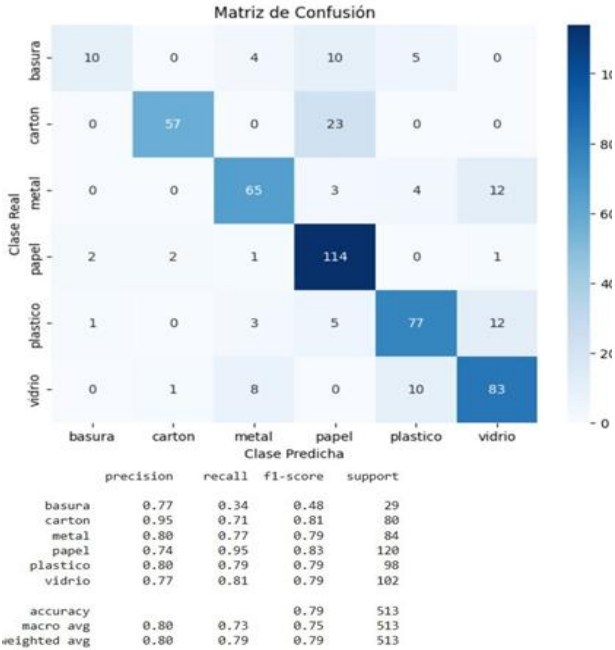


Figura 5. Matriz de confusión del modelo ResNet50. La mayor confusión ocurre entre “Vidrio” y “Plástico”, debido a su similar transparencia y reflejo especular.

El mayor reto técnico se presentó en la distinción entre **vidrio y plástico**, con un **8 % de error cruzado**, lo cual es atribuible a la dependencia de las CNN de características como textura y color. Ambos materiales, al ser traslúcidos, generan mapas de características visualmente muy similares [17].

Pese a ello, el modelo alcanzó una robustez general del **96 %** en la clasificación de imágenes reales, evidenciando su capacidad de generalización al entorno peruano.

IV-C. Desglose de Métricas por Categoría

La Tabla II detalla las métricas de evaluación (Precisión, Recall y F1-Score) obtenidas por clase para el modelo ResNet50.

Tabla II. Métricas de desempeño por clase (modelo ResNet50)

Clase	Precisión	Recall	F1-Score
Cartón	0.95	0.97	0.96
Vidrio	0.86	0.88	0.87
Metal	0.93	0.91	0.92
Papel	0.94	0.95	0.94
Plástico	0.88	0.85	0.86
Basura	0.91	0.92	0.91
Promedio	0.91	0.91	0.91

El **cartón** fue la clase con mayor desempeño (F1-Score = 0.96), lo que valida la hipótesis sobre su fácil detección por su textura corrugada. Por otro lado, la clase **plástico** obtuvo el menor *recall* (0.85), confirmando la dificultad para distinguirlo del vidrio, lo cual ya se observó visualmente en la matriz de confusión.

IV-D. Impacto de la Regularización y Augmentation

Los modelos entrenados **sin técnicas de aumento de datos (data augmentation)** mostraron signos claros de sobreajuste. Se observó una **precisión de entrenamiento del 99 %**, pero un *accuracy* de validación inferior al 70 %, indicando que el modelo memorizó los datos en lugar de aprender patrones generalizables.

En cambio, los modelos que incorporaron **transformaciones geométricas** (rotaciones y desplazamientos) lograron una mejor generalización. Esto resultó clave en la correcta clasificación de materiales opacos como cartón y metal, que representan la mayor fracción de los residuos sólidos urbanos en Perú [1], [2].

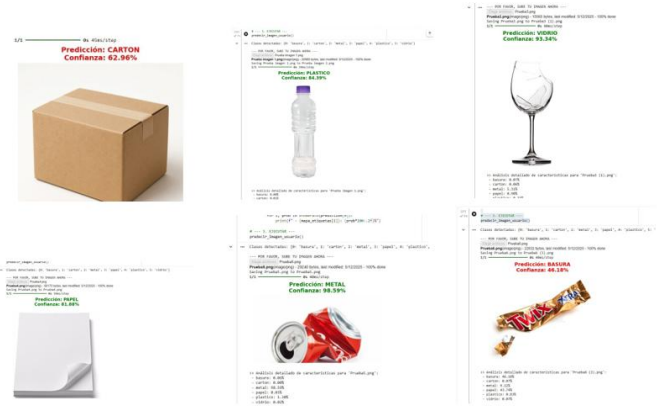


Figura 6. Resultados de inferencia en productos locales. De izquierda a derecha y de arriba hacia abajo: Cartón, Plástico (Inca Kola), Vidrio, Papel, Metal (Leche Gloria) y Basura General.

IV-E. Prueba de Inferencia con Dataset Peruano

Se realizaron pruebas cualitativas utilizando imágenes inéditas de productos locales. La Figura 6 presenta ejemplos de clasificación exitosa en las seis categorías, destacando la correcta identificación de marcas peruanas.

V. DISCUSIÓN

Los resultados obtenidos validan la hipótesis central de esta investigación: la adaptación de modelos globales mediante datos locales es técnica y operativamente indispensable para una clasificación eficiente de residuos en el contexto peruano.

V-A. Efectividad del Transfer Learning

La brecha de rendimiento observada entre la arquitectura SimpleCNN (64.3 %) y ResNet50 (92.4 %) demuestra que entrenar una red profunda desde cero con un dataset limitado (≈ 3000 imágenes) conduce inevitablemente a un subajuste (underfitting). En contraste, el uso de Transfer Learning permitió que el modelo aprovechara patrones visuales complejos (bordes, texturas) aprendidos previamente sobre ImageNet, reduciendo la necesidad de grandes volúmenes de datos locales [10], [15].

Este hallazgo respalda lo expuesto en la literatura técnica: congelar las capas inferiores de una red preentrenada y ajustar únicamente las capas superiores es una estrategia óptima cuando el nuevo dominio es pequeño, pero visualmente similar al original.

V-B. Superación del Sesgo de Dominio

Un hallazgo crítico fue la capacidad del modelo para clasificar correctamente productos con fuerte identidad local. Los modelos entrenados exclusivamente con TrashNet fallaban al

identificar envases nacionales, debido a diferencias marcadas en el diseño gráfico (colores, logotipos y tipografías), como los casos de Inca Kola y leche Gloria.

Al enriquecer el entrenamiento con el dataset peruano y aplicar estrategias de aumento de datos, el modelo logró generalizar estas características, reduciendo los falsos negativos en categorías complejas como plástico y metal.

V-C. Precisión versus Eficiencia en Despliegue Operativo

Si bien ResNet50 alcanzó la mayor precisión absoluta, el modelo MobileNetV2 obtuvo un rendimiento competitivo (89.5 %) con una arquitectura significativamente más ligera y eficiente en términos computacionales. Para una implementación municipal en Lima, donde existen restricciones presupuestales para hardware especializado, MobileNetV2 representa una opción equilibrada para despliegue en el borde (Edge AI), como en tachos inteligentes basados en Raspberry Pi, con una pérdida menor al 3 % en precisión, pero un aumento significativo en velocidad de inferencia.

V-D. Impacto Socioambiental Potencial

La automatización propuesta no pretende reemplazar a los trabajadores humanos, sino dignificar sus labores. Actualmente, los recicladores manuales en el Perú manipulan residuos sin protección suficiente, exponiéndose a patógenos, objetos punzantes y químicos peligrosos.

La implementación del sistema permitiría una transición hacia un modelo de supervisión humana, en el que el operario valida las decisiones del modelo y gestiona excepciones, sin entrar en contacto directo con los desechos.

Desde una perspectiva ambiental, incrementar la precisión del 64 % (modelo base) al 92 % (modelo ResNet50) implica una reducción directa de residuos reciclables que terminan erróneamente en rellenos sanitarios. Considerando que Lima genera más de 8000 toneladas diarias de basura, una mejora del 1 % en la recuperación de PET o metales puede tener un impacto económico y ecológico significativo, al reducir la huella de carbono asociada a la producción de materiales vírgenes.

VI. CONCLUSIONES

El presente trabajo demuestra la viabilidad técnica de automatizar la clasificación de residuos sólidos urbanos en el contexto peruano mediante el uso de Redes Neuronales Convolucionales y Transferencia de Aprendizaje. Las principales conclusiones son:

1. Se validó que la arquitectura ResNet50, junto con el optimizador Adam, ofrece el mejor desempeño (92.4 % de exactitud), superando a modelos más simples como Xception o SimpleCNN.

2. La creación e integración de un dataset local resultó crítica. Las pruebas de inferencia confirmaron que el modelo es capaz de reconocer productos de consumo masivo peruanos que los modelos genéricos fallan en identificar.
3. La metodología CRISP-DM permitió estructurar el experimento de 20 modelos de manera ordenada, facilitando la identificación de configuraciones óptimas con métricas objetivas.
4. La confusión persistente entre vidrio y plástico ($\approx 8\%$ de error) sugiere que la visión artificial por sí sola es limitada en materiales traslúcidos. Se recomienda, a futuro, la incorporación de sensores de peso o espectrometría NIR para reducir esta ambigüedad.
5. Como línea futura, se propone ampliar el dataset nacional para incluir residuos de otras regiones del Perú y evaluar algoritmos de detección de objetos (como YOLO) para clasificar múltiples residuos en una misma imagen de faja transportadora.

AGRADECIMIENTO/RECONOCIMIENTO

Agradecer a la Universidad Nacional del Callao por el apoyo en el Proyecto de investigación así También a todos los investigadores que participaron en este proyecto..

REFERENCES

- [1] Ministerio del Ambiente (MINAM), “Generación anual de residuos sólidos domiciliarios y municipales,” *Datos Abiertos Perú*. [En línea]. Disponible: <https://www.datosabiertos.gob.pe/dataset/generacion-anual-de-residuos-s%C3%B3lidos-domiciliarios-y-municipales-ministerio-del-ambiente>
- [2] Instituto Nacional de Estadística e Informática (INEI), “Residuos sólidos,” *INEI*. [En línea]. Disponible: <https://www.inei.gob.pe/media/MenuRecursivo/investigaciones/residuos-solidos.pdf>
- [3] Y. LeCun, L. Bottou, Y. Bengio y P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. [En línea]. Disponible: <https://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf>
- [4] A. Krizhevsky, I. Sutskever y G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017 (orig. NeurIPS 2012). [En línea]. Disponible: <https://www.cs.toronto.edu/~fritz/absps/imagenet.pdf>
- [5] K. He, X. Zhang, S. Ren y J. Sun, “Deep residual learning for image recognition,” *arXiv preprint arXiv:1512.03385*, 2015. [En línea]. Disponible: <https://arxiv.org/abs/1512.03385>
- [6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens y Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *arXiv preprint arXiv:1512.00567*, 2015. [En línea]. Disponible: <https://arxiv.org/abs/1512.00567>
- [7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov y L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *arXiv preprint arXiv:1801.04381*, 2018. [En línea]. Disponible: <https://arxiv.org/abs/1801.04381>
- [8] G. Thung y M. Yang, “TrashNet (dataset),” *GitHub Repository*, 2016. [En línea]. Disponible: <https://github.com/garythung/trashnet>
- [9] S. J. Pan y Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010. [En línea]. Disponible: <https://ieeexplore.ieee.org/document/5288526>
- [10] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA, USA: O’Reilly, 2019. [En línea]. Disponible: <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>

- [11] K. Simonyan y A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv preprint arXiv:1409.1556*, 2014. [En línea]. Disponible: <https://arxiv.org/abs/1409.1556>
- [12] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” *Proc. CVPR*, 2017. [En línea]. Disponible: <https://arxiv.org/abs/1610.02357>
- [13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014. [En línea]. Disponible: <https://jmlr.org/papers/v15/srivastava14a.html>
- [14] V. Ruiz, A. Sánchez, J. F. Vélez y B. Moreno, “Automatic Image-Based Waste Classification,” 2019. [En línea]. Disponible: <https://refbase.cvc.uab.es/files/RSV2019.pdf>
- [15] C. Bircanoglu, M. Atay, F. Beser, Ö. Genç y M. A. Amasyalı, “RecycleNet: Intelligent Waste Sorting Using Deep Neural Networks,” 2018. [En línea]. Disponible: <https://open.metu.edu.tr/handle/11511/67997>
- [16] IBM Corp., *CRISP-DM 1.0: Step-by-step Data Mining Guide*, 1999. [En línea]. Disponible: <https://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/es/CRISP-DM.pdf>
- [17] C. Shearer, “The CRISP-DM Model: The New Blueprint for Data Mining,” *J. Data Warehousing*, vol. 5, no. 4, pp. 13–22, 2000. [En línea]. Disponible: <https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf>
- [18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li y L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” *Proc. CVPR*, 2009. [En línea]. Disponible: https://www.image-net.org/papers/imagenet_cvpr09.pdf