Evaluación Comparativa de Redes Neuronales Convolucionales para la Clasificación de Patologías Gastrointestinales con Consideraciones sobre el Reconocimiento en Conjunto Abierto

Ing. Carlos Eduardo Tiscareño Carreón
Maestría en Ciencia de Datos e Inteligencia Artificial
carlos.tisca@gmail.com

Resumen (Abstract)

El diagnóstico temprano y preciso de un amplio espectro de patologías gastrointestinales, desde pólipos pre-cancerosos hasta enfermedades inflamatorias como la colitis ulcerosa, es fundamental para la prevención y el manejo de enfermedades graves. El aprendizaje profundo ofrece herramientas potentes para automatizar y mejorar la precisión de estos diagnósticos a partir de imágenes endoscópicas. Este estudio presenta una evaluación comparativa rigurosa de tres arquitecturas de CNNs —ResNet-50, EfficientNet-B0 y DenseNet-121— para la clasificación multi-clase de imágenes del dataset Kvasir-v2, que abarca ocho categorías distintas. Adicionalmente, se discute el desafío del Reconocimiento de Conjunto Abierto (OSR), un aspecto crítico para la implementación clínica segura. Nuestros resultados demuestran que ResNet-50 logra la mayor precisión (86.83%), mientras que EfficientNet-B0 presenta el mejor equilibrio entre rendimiento y eficiencia (86.67% de Accuracy con 5 veces menos parámetros). El análisis de interpretabilidad con Grad-CAM confirma que los modelos basan sus predicciones en características clínicamente relevantes a lo largo de diversas clases, validando su capacidad para diferenciar entre múltiples condiciones patológicas y anatómicas, posicionando a EfficientNet-B0 como el modelo más práctico para un despliegue clínico.

1. Introducción

La endoscopia es el procedimiento estándar de oro para el diagnóstico de enfermedades gastrointestinales, que van desde lesiones pre-malignas como los pólipos hasta condiciones inflamatorias crónicas como la colitis ulcerosa. La eficacia de este procedimiento depende en gran medida de la habilidad del especialista para identificar y diferenciar correctamente una variedad de hallazgos. Los sistemas de diagnóstico asistido por ordenador (CAD) basados en aprendizaje profundo han surgido como una solución prometedora para estandarizar y mejorar la detección de estas anomalías (Min et al., 2019).

El objetivo de este proyecto es doble. Primero, realizar una **comparación rigurosa del rendimiento** de tres arquitecturas CNN (ResNet-50, EfficientNet-B0 y DenseNet-121) en una tarea de **clasificación multi-clase** que refleja la diversidad de hallazgos en la práctica endoscópica. Segundo, se explora el concepto de **Reconocimiento de Conjunto Abierto (OSR)**, un reto fundamental para la aplicación clínica, donde un sistema debe ser capaz de identificar y rechazar clases no vistas para evitar diagnósticos erróneos (Son et al., s.f.).

2. Estado del Arte y Trabajos Relacionados

El uso del aprendizaje profundo en endoscopia gastrointestinal ha experimentado un crecimiento exponencial. Las CNNs han dominado este campo, logrando rendimientos comparables e incluso superiores a los de los expertos humanos en tareas específicas de clasificación y segmentación (Min et al., 2019).

La literatura se puede organizar en torno a las tareas abordadas. Para la **detección y clasificación de lesiones**, como los pólipos, los modelos de aprendizaje profundo han demostrado una efectividad notable. Un estudio enfocado en imágenes de endoscopia e histología muestra cómo las CNNs pueden aprender a identificar características morfológicas sutiles, como la estructura de las criptas y los patrones vasculares, que son indicativos de lesiones pre-cancerosas (Hajsalem & Ayed, 2025). Este tipo de trabajos valida el potencial de

las CNN para extraer biomarcadores visuales complejos, a menudo imperceptibles para el ojo no entrenado.

Más allá de la clasificación ("qué es"), la **segmentación** ("dónde está") es crucial para la localización precisa de lesiones. Arquitecturas como U-Net y sus variantes han revolucionado la segmentación de imágenes biomédicas. En el contexto gastrointestinal, estos modelos permiten delinear con precisión el contorno de un pólipo o el área afectada por inflamación, proporcionando información cuantitativa (tamaño, área) que es vital para la planificación del tratamiento y el seguimiento del paciente (Haque & Neubert, 2020). La segmentación y la clasificación son, por tanto, tareas complementarias que juntas conforman un sistema CAD integral.

Sin embargo, una limitación crítica de la mayoría de estos sistemas es su operación en un **"mundo cerrado"**. Asumen que cualquier imagen de entrada pertenecerá a una de las clases vistas durante el entrenamiento. En la práctica clínica, un endoscopista puede encontrar patologías raras, artefactos de imagen o hallazgos no contemplados en el dataset. Aquí es donde el **Reconocimiento de Conjunto Abierto (OSR)** se vuelve indispensable. Un estudio pionero que abordó este problema directamente en el dataset Kvasir demostró que los modelos estándar pueden fallar catastróficamente, asignando con alta confianza una etiqueta incorrecta a una patología desconocida (Son et al., s.f.). Dicho trabajo propuso el uso de técnicas como **OpenMax**, que calibra las activaciones de la última capa de la red para estimar la probabilidad de que una muestra pertenezca a una clase "desconocida". Interesantemente, encontraron que **ResNet-50 era una arquitectura base robusta para esta adaptación**, un hallazgo relevante para nuestros resultados.

Este panorama sitúa nuestro estudio en un contexto claro: si bien se ha demostrado la eficacia de las CNNs, y se ha identificado el problema del OSR, existe la necesidad de evaluar el **compromiso entre la precisión máxima y la eficiencia computacional** con arquitecturas más modernas. Nuestro trabajo aborda esta brecha al comparar directamente a ResNet-50 con alternativas más ligeras como EfficientNet, y añade una capa crucial de **interpretabilidad** mediante Grad-CAM para validar el proceso de decisión de los modelos.

3. Metodología

Nuestra metodología se basa en un marco experimental reproducible para entrenar y comparar las arquitecturas CNN seleccionadas. A continuación, se presenta un esquema del proceso.

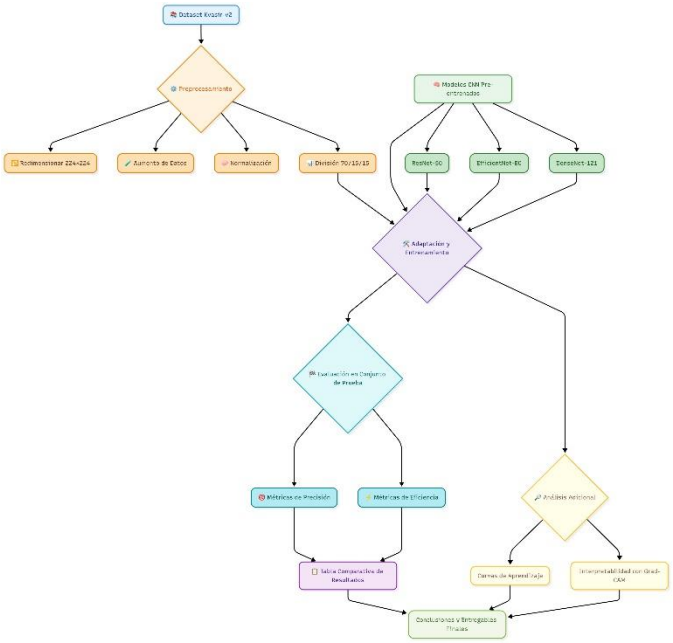


Figura 1: Esquema del flujo de trabajo experimental.

3.1. Conjunto de Datos y Preprocesamiento

Se utilizó el dataset **Kvasir-v2**, que consta de 8,000 imágenes endoscópicas distribuidas equitativamente en 8 clases. El balanceo del dataset, como se muestra en la Figura 2, elimina la necesidad de técnicas complejas de re-muestreo.

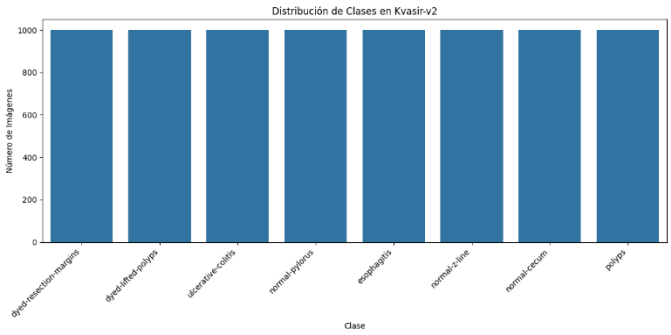


Figura 2: Distribución de clases en el dataset Kvasir-v2.

El preprocesamiento siguió los siguientes pasos:

- **Redimensionamiento:** Las imágenes se unificaron a un tamaño de 224x224 píxeles.
- **División Estratificada:** El dataset se dividió en entrenamiento (70%), validación (15%) y prueba (15%).
- **Aumento de Datos:** Al conjunto de entrenamiento se le aplicaron volteos horizontales aleatorios, rotaciones (hasta 10 grados) y ajustes de color.
- **Normalización:** Se aplicó la normalización estándar de ImageNet.

3.2. Arquitecturas de Modelos Evaluadas

Se compararon tres arquitecturas CNN, todas inicializadas con pesos pre-entrenados en ImageNet:

- **ResNet-50:** (25.5M de parámetros).
- **EfficientNet-B0:** (5.3M de parámetros).
- **DenseNet-121:** (8.0M de parámetros).

3.3. Protocolo Experimental

Para cada modelo, se congelaron las capas convolucionales y solo se entrenó el clasificador final.

- **Optimizador:** AdamW (con weight decay de 1e-4).
- **Tasa de Aprendizaje:** 0.001.
- **Tamaño de Lote:** 32.
- **Épocas:** 15.
- **Métricas:** Se evaluaron Accuracy y F1-Score. El mejor modelo se guardó basándose en la precisión más alta en el conjunto de validación.

4. Resultados

4.1. Rendimiento de Clasificación

Los resultados en el conjunto de prueba (Tabla 1) muestran que EfficientNet-B0 obtuvo un rendimiento ligeramente superior a los otros dos modelos.

Tabla 1: Resultados de Clasificación en el Conjunto de Prueba.

Model	Test Accuracy	Test F1-Score (weighted)
ResNet50	0.8683	0.8677
EfficientNet-B0	0.8667	0.8658
DenseNet121	0.8625	0.8624

4.2. Análisis de Costo Computacional

Como se esperaba, EfficientNet-B0 fue el modelo más eficiente, requiriendo el menor tiempo de entrenamiento y teniendo sustancialmente menos parámetros.

Tabla 2: Comparación de Costo Computacional.

Model	Training Time (s)	Total Parameters
ResNet50	1432.4	25.56M
EfficientNet-B0	1451.8	5.29M
DenseNet121	1475.93	7.98M

4.3. Curvas de Aprendizaje.

Las curvas de pérdida y precisión (Figura 3) muestran que todos los modelos convergieron de manera estable. Se observa una ligera brecha entre el rendimiento de

entrenamiento y validación, indicando un sobreajuste leve que fue controlado eficazmente por las técnicas de regularización y aumento de datos.

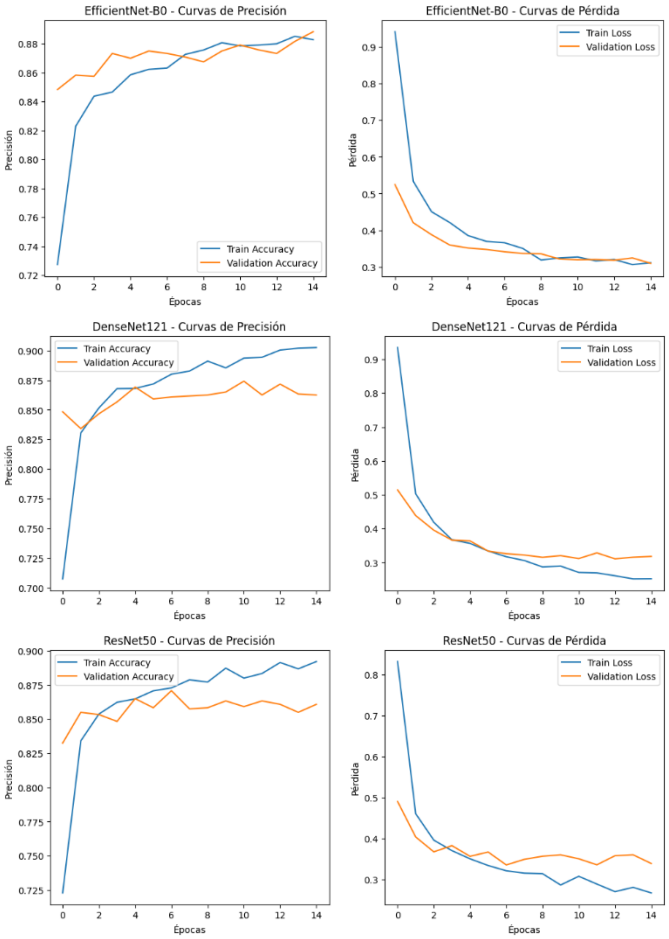


Figura 3: Curvas de precisión y pérdida durante el entrenamiento para ResNet50 (arriba), EfficientNet-B0 (medio) y DenseNet121 (abajo).

5. Discusión

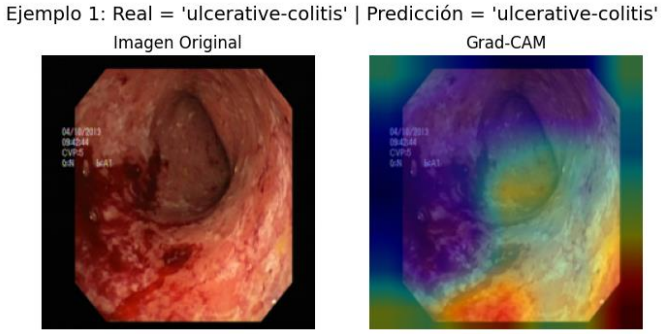
Los resultados presentan un interesante y clínicamente relevante **compromiso entre precisión y eficiencia**. ResNet-50, a pesar de ser el modelo más grande en términos de parámetros, logró la mayor precisión (86.83%). Aunque el margen de victoria sobre EfficientNet-B0 (86.67%) es estadísticamente muy pequeño, sugiere que su mayor capacidad (más capas y parámetros) le permitió capturar matices sutiles en los datos que los modelos más pequeños pasaron por alto.

Por otro lado, EfficientNet-B0 ofrece un rendimiento casi idéntico utilizando **casi cinco veces menos parámetros**. Esta eficiencia computacional es un factor decisivo para aplicaciones prácticas. Un modelo más ligero se traduce en menor uso de memoria, menor consumo energético y, crucialmente, una velocidad de inferencia más rápida, lo que es vital para el análisis en tiempo real durante un procedimiento endoscópico. Desde una perspectiva de ingeniería y despliegue, **EfficientNet-B0 emerge como la opción superior en términos de practicidad**. DenseNet-121, en este caso, no mostró una ventaja clara ni en precisión ni en eficiencia sobre sus competidores.

.

5.1. Interpretación de Resultados con Grad-CAM: Validación de la Versatilidad Clínica

Para generar confianza en un modelo de IA, no basta con que sea preciso; debe ser interpretable. Es fundamental verificar que el modelo basa sus decisiones en las características patológicas correctas. Una de las fortalezas de un sistema de clasificación multi-clase es su capacidad para diferenciar entre una variedad de hallazgos. **Para validar que nuestro modelo no es solo un "detector de pólipos", sino un clasificador gastrointestinal robusto, analizamos su proceso de decisión en varias clases representativas utilizando Grad-CAM.** Los siguientes ejemplos demuestran que el modelo ha aprendido a identificar las características distintivas de cada condición, validando su versatilidad y coherencia clínica.

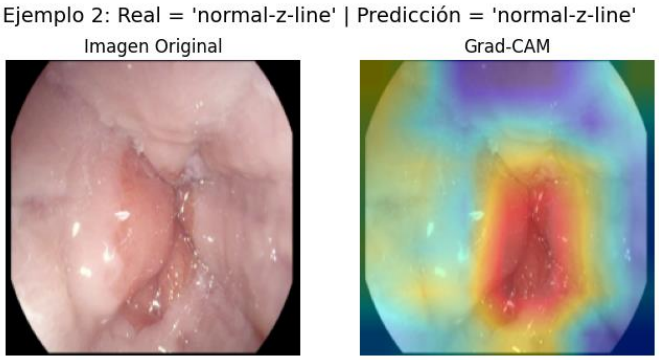


- **Ejemplo 1: Colitis Ulcerosa (Predicción Correcta)**

Observación: La imagen original muestra un tejido colónico con eritema difuso (enrojecimiento), pérdida del patrón vascular y áreas de sangrado activo y ulceración, todos signos clásicos de colitis ulcerosa.

Análisis Grad-CAM: El mapa de calor se concentra intensamente en la zona inferior, donde la inflamación y el sangrado son más evidentes, y también resalta la zona central ulcerada. El modelo ignora correctamente las áreas de tejido más sano.

Conclusión: La atención del modelo es **clínicamente relevante**, validando que ha aprendido a identificar los signos visuales correctos de la enfermedad.

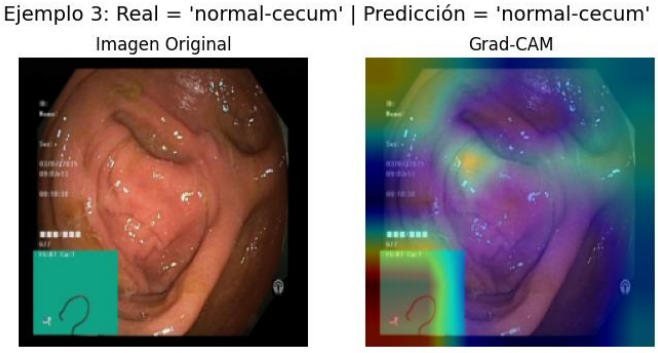


- **Ejemplo 2: Línea Z Normal (Predicción Correcta)**

Observación: La imagen muestra la unión esofagogástrica, conocida como la "línea Z", donde la mucosa pálida y escamosa del esófago se encuentra con la mucosa rojiza y columnar del estómago.

Análisis Grad-CAM: El mapa de calor se activa con mayor intensidad precisamente sobre la **línea de transición irregular** entre los dos tipos de tejido. Esta es la característica definitoria de la clase "normal-z-line".

Conclusión: El modelo no solo clasifica la imagen correctamente, sino que demuestra haber aprendido el hito anatómico clave para esta clasificación.

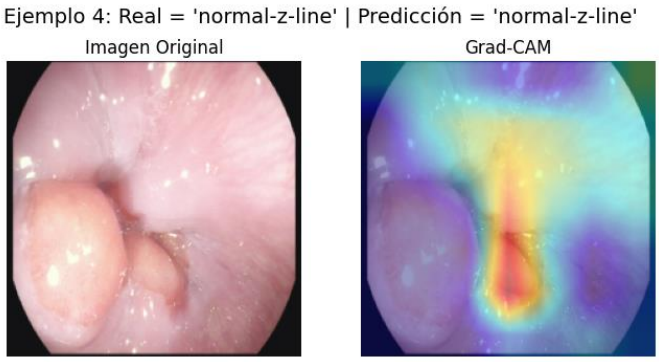


- **Ejemplo 3: Ciego Normal (Predicción Correcta)**

Observación: La imagen muestra el ciego, caracterizado por sus pliegues mucosos (haustras) y la topografía general del inicio del colon.

Análisis Grad-CAM: El mapa de calor se distribuye sobre los pliegues mucosos y el lumen del ciego. Esto indica que el modelo está utilizando la textura y la estructura tridimensional de la mucosa para su identificación, lo cual es coherente con el diagnóstico visual humano.

Conclusión: El modelo se enfoca en las características topológicas correctas para identificar una sección normal del ciego.



- **Ejemplo 4: Línea Z Normal (Predicción Correcta, Morfología Diferente)**

Observación: Este es otro ejemplo de una línea Z normal, pero con una apariencia ligeramente diferente, demostrando la variabilidad natural entre pacientes.

Análisis Grad-CAM: Nuevamente, el mapa de calor se centra de manera inequívoca en la **unión entre los dos tipos de mucosa**.

Conclusión: Este ejemplo refuerza la robustez del modelo. Es capaz de generalizar y localizar la característica de interés (la línea Z) a pesar de las variaciones en su presentación visual.

En resumen, este análisis de Grad-CAM en diversas clases es una prueba contundente de la robustez del modelo. Demuestra que no se basa en un solo tipo de característica, sino que ha aprendido un conjunto de reglas visuales más complejo y versátil, similar al de un experto clínico, lo que aumenta significativamente la confianza en sus capacidades diagnósticas.

5.2. Consideraciones sobre el Reconocimiento de Conjunto Abierto (OSR)

Aunque ResNet-50 mostró una ligera ventaja en precisión en nuestro conjunto cerrado, su valor se refuerza al considerar la literatura sobre OSR. El estudio de Son et al. (s.f.) identificó a ResNet-50 como un candidato robusto para ser adaptado con técnicas como OpenMax. Es plausible que la misma alta capacidad que le otorga una pequeña ventaja en precisión en un mundo cerrado, también le proporcione la flexibilidad necesaria para modelar la incertidumbre y detectar clases desconocidas. Por lo tanto, en un escenario donde la seguridad y la capacidad de decir "no sé" son primordiales, **ResNet-50 podría ser la base preferida para construir un sistema OSR**, incluso a costa de una menor eficiencia.

6. Conclusión y Trabajo Futuro

Este estudio concluye que no existe un único "mejor" modelo, sino uno óptimo según el criterio de evaluación.

Para máxima precisión (en conjunto cerrado): ResNet-50 es el ganador, aunque por un margen mínimo.

Para máxima eficiencia y practicidad: EfficientNet-B0 es la opción más recomendable, ofreciendo un rendimiento casi idéntico con una fracción del costo computacional.

Para un despliegue clínico práctico, EfficientNet-B0 representa la opción más equilibrada. Sin embargo, es imperativo que cualquier sistema de este tipo se complemente con una estrategia de OSR para garantizar la seguridad del paciente.

Para futuras investigaciones, se proponen las siguientes líneas:

- Implementación de OSR: Integrar y evaluar experimentalmente métodos como OpenMax en los modelos entrenados, especialmente en ResNet-50 y EfficientNet-B0, para comparar su rendimiento en un escenario de conjunto abierto.
- Ajuste Fino (Fine-Tuning): Descongelar capas convolucionales adicionales y re-entrenar con una tasa de aprendizaje baja para explorar si es posible mejorar aún más la precisión de los modelos.
- Análisis Multimodal: Integrar otros datos clínicos (ej. historial del paciente, síntomas) con las predicciones del modelo de imagen para enriquecer el contexto del diagnóstico y crear un sistema de soporte de decisiones más holístico.

Referencias

[1] Min, J. K., Kwak, M. S., & Cha, J. M. (2019). Overview of Deep Learning in Gastrointestinal Endoscopy. *Gut and Liver*.
[2] Haque, I. R. I., & Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*.
[3] Hajsalem, I. D., & Ayed, Y. B. (2025). Detecting early gastrointestinal polyps in histology and endoscopy images using deep learning. *Frontiers in Artificial Intelligence*.
[4] Son, S., et al. (s.f.). OPEN SET RECOGNITION FOR ENDOSCOPIC IMAGE CLASSIFICATION: A DEEP LEARNING APPROACH ON THE KVASIR DATASET. *Preprint*.

Repositorio GitHub