



Revista EIA  
ISSN 1794-1237  
e-ISSN 2463-0950  
Año XIX/ Volumen 22/ Edición N.43  
Enero - junio 2025  
Reia4317 pp. 1-28

Publicación científica semestral  
Universidad EIA, Envigado, Colombia

## PARA CITAR ESTE ARTÍCULO / TO REFERENCE THIS ARTICLE /

García-Arias, S. A. y Velandia, F.  
Modelos de redes neuronales  
convolucionales como herramienta  
para automatizar la clasificación de  
rocas

Revista EIA, 22(43), Reia4317 pp. 1-28  
<https://doi.org/10.24050/reia.v22i43.1813>

 Autor de correspondencia:

García-Arias, S. A.  
Geólogo y Magister en Geofísica.  
Correo electrónico:  
[sergio.garcia.910815@gmail.com](mailto:sergio.garcia.910815@gmail.com)

**Recibido:** 01-08-2024

**Aceptado:** 10-12-2024

**Disponible online:** 01-01-2025

# Modelos de redes neuronales convolucionales como herramienta para automatizar la clasificación de rocas

 SERGIO A. GARCÍA-ARIAS<sup>1</sup>

FRANCISCO VELANDIA<sup>1</sup>

1. Escuela de Geología, Universidad Industrial de Santander, Colombia

## Resumen

En geología, la clasificación de las rocas ígneas, sedimentarias y metamórficas en distintas escalas es esencial para el estudio de una región. Los avances tecnológicos han permitido incorporar herramientas que facilitan estos procesos de observación y análisis, beneficiando tanto la formación de profesionales como el acceso al conocimiento geológico para todos los interesados. Como parte de las herramientas de machine learning, las Convolutional Neural Networks se destacan para identificar y priorizar características de objetos en una imagen. Entre los modelos relevantes de este grupo se encuentra You Only Look Once (YOLO), los cuales sirven como base para el reentrenamiento de modelos más especializados. El objetivo de este trabajo es la creación de modelos reentrenados basados en YOLO para automatizar la identificación y clasificación de diferentes tipos de rocas.

Se realizó el reentrenamiento de dos arquitecturas YOLO, YOLOv8n y YOLOv8x, utilizando imágenes de muestras de rocas provenientes de repositorios en línea y del catálogo de la Escuela de Geología de la Universidad Industrial de Santander. Estas imágenes fueron verificadas y validadas bajo criterios geológicos. En cada etapa de entrenamiento se realizaron cambios a los catálogos de imágenes y parámetros de entrenamiento. La validación de los modelos entrenados se efectuó con muestras de roca de la Escuela de Geología que no fueron utilizadas en la etapa de entrenamiento. Para cuantificar la calidad de los modelos se utilizó el cálculo de la matriz de confusión normalizada en función de las etiquetas reales, obteniendo así el porcentaje de aciertos para cada tipo de roca de manera proporcional.

Un total de once modelos reentrenados se obtuvieron en las diferentes etapas de entrenamiento. Dos modelos lograron una mayor capacidad de generalización y obtuvieron valores superiores al 30% en la diagonal de las matrices de confusión. Otros modelos alcanzaron un valor superior al 80% para la identificación de rocas ígneas. Para la identificación de rocas sedimentarias, un modelo logró un acierto del 93%, mientras que, para las rocas metamórficas, otro obtuvo un acierto del 61%. Estos modelos pueden servir como base para desarrollar otros más complejos en proyectos específicos. Considerando que, aunque los modelos no fueron entrenados con imágenes de afloramientos, es posible estimar el tipo de roca al que pertenecen y se resalta su capacidad para extraer características básicas.

**Palabras clave:** yolo; aprendizaje profundo; modelo de clasificación; segmentación; transfer learning; geología; petrografía; reconocimiento de patrones; muestras de roca; procesamiento de imágenes.

---

## Convolutional Neural Network models as a tool for automating rock classification

### Abstract

Classifying rocks into igneous, sedimentary, and metamorphic types at different scales is essential for studying a region. Technological advancements facilitate these observation and analysis processes, benefiting both the training of professionals and the accessibility of geological knowledge for all interested parties. As one of the machine learning tools, Convolutional Neural Networks (CNNs) excel in identifying and prioritizing object features in an image. Among the relevant models in this group is You Only Look Once (YOLO), which serves as a foundation for retraining more specialized models. This work aims to create retrained models based on YOLO to automate identifying and classifying different types of rocks.

Two YOLO architectures, YOLOv8n and YOLOv8x were retrained using rock sample images from online repositories and the catalog of the School of Geology of the Universidad Industrial de Santander. These images were verified and validated based on geological criteria. Throughout each training stage, changes were made to the image catalogs and training parameters. The validation of the trained models was performed using rock samples from the School of Geology that were not used in the training. To quantify the quality of the models, the calculation of the normalized confusion matrix was used based on the actual labels, thereby obtaining the percentage of correct identifications for each rock type proportionally.

A total of eleven retrained models were obtained during the different training stages. Two models achieved a higher generalization capacity, obtained values greater than 30% on the diagonal of the confusion matrices. Other two models reached values greater than 80% for identifying igneous rocks. For sedimentary rocks, one model achieved an accuracy of 93%, while for metamorphic rocks, another obtained an accuracy of 61%. These models can serve as a foundation for developing more complex models for specific projects. Even though they were not trained with outcrop images, it is possible to estimate the type of rock they belong to, highlighting the capability of these models to extract basic features.

**Keywords:** yolo; deep learning; classification model; segmentation; transfer learning; geology; petrography; pattern recognition; rock samples; image processing.

## 1. Introducción

Los minerales se consideran sólidos inorgánicos naturales con una estructura interna ordenada y una composición química específica. Una roca es una masa sólida de materia mineral o similar que se presenta naturalmente en nuestro planeta. Algunas rocas están compuestas casi por completo de un solo mineral. De manera general, las rocas se categorizan en tres tipos principales: ígneas, sedimentarias y metamórficas. Las rocas ígneas se forman a partir del enfriamiento y solidificación del magma o roca fundida. Las rocas sedimentarias se forman cuando los materiales inconsolidados y transportados se acumulan en depresiones o a nivel de la superficie de la Tierra; los sedimentos se generan a partir de rocas preexistentes mediante procesos de meteorización y erosión. Las rocas metamórficas se generan a partir de rocas ígneas, sedimentarias u otras metamórficas, denominadas “rocas madre”, que se transforman por efecto de altas temperaturas y presiones en la profundidad de la corteza terrestre y el manto superior (Tarbuck & Lutgens, 2005).

En geología, la clasificación de las rocas es un proceso necesario para el estudio de una región y conlleva trabajo de observación en distintas escalas, desde la regional, mediante el análisis del paisaje, hasta la escala meso en afloramientos y micro con secciones delgadas (petrografía). Usualmente, la discriminación de los tres

tipos principales de rocas se consigue a simple vista después de un entrenamiento del observador, y luego con el uso de la lupa para diferenciar la textura y composición de muestras de mano y afloramientos. La clasificación litológica más específica se logra mediante el uso de microscopios para diferenciar los tipos de minerales que las componen.

Actualmente, con el avance de la tecnología, es necesario incorporar nuevas herramientas que facilitan el proceso de observación y análisis en Geología, no solamente como apoyo en el entrenamiento de los nuevos profesionales, sino para permitir el acceso al conocimiento geológico a todos los interesados, especialmente es esta época de trabajo transdisciplinario, además para satisfacer la curiosidad de las personas que cada vez se involucran y sensibilizan más con el planeta. Desde este punto de vista, consideramos importante el uso de equipos disponibles en las instituciones educativas o de investigación, tal que soporten aplicaciones para generar modelos que nos permiten una clasificación litológica preliminar y rápida.

Entre los métodos de aprendizaje automático se encuentran las redes neuronales (Hastie et al., 2009; MacKay, 2003; Rohwer et al., 1994; Rojas, 1996) como las de mayor interés para ser utilizados en las geociencias, en la identificación de arcillas (Gómez et al., 2013), generación de mapas predictivos (García-Arias et al., 2024; Goyes-Peñañiel & Hernández-Rojas, 2021), clasificación litológica (Cracknell, 2014), clasificación de rocas en secciones delgadas (Młynarczyk et al., 2013), aplicaciones geofísicas (Dramsch, 2020; van der Baan & Jutten, 2000) entre otros. Con el tiempo, se han desarrollado modelos más complejos de redes neuronales, como las redes neuronales convolucionales (Convolutional Neural Networks, CNN), específicamente para la clasificación y segmentación de imágenes (se explica en el siguiente párrafo). Bajo esta consideración, las CNN se pueden considerar como ideales para la clasificación de rocas sedimentarias, ígneas y metamórficas, ya que pueden identificar texturas, formas y composiciones características de cada tipo de roca.

Una CNN es un enfoque de aprendizaje profundo para asignar prioridad a varios aspectos u objetos en una imagen, y al mismo tiempo identificarlos (Hussain et al., 2019). La detección de

características en una CNN se realiza en cada capa considerando los datos de entrenamiento. Debido al mecanismo de reparto de peso, las neuronas en la misma superficie de mapeo de características, tienen el mismo peso (Xin & Wang, 2019). Algunos aspectos importantes por considerar en las CNN incluyen la selección de arquitecturas, lo que influye en la cantidad de parámetros del modelo, además, es crucial determinar el efecto de la escala del conjunto de datos y el contexto de las imágenes de entrenamiento (Shin et al., 2016). La segmentación de imágenes tiene que ver con dividir una imagen digital en grupos de píxeles, o segmentos de imagen, con el objetivo de facilitar la detección de objetos y otras tareas relacionadas. Este proceso permite identificar y diferenciar claramente distintas áreas dentro de una imagen, lo que puede tener un impacto significativo en el rendimiento y la capacidad de generalización de los modelos de redes neuronales convolucionales (CNN) en tareas como la clasificación de rocas.

En las geociencias se han utilizado las CNN para la clasificación de rocas en distintas escalas, desde el nivel macroscópico (Liu et al., 2020; Patro et al., 2022), hasta el microscópico en sección delgada (Cheng & Guo, 2017; Ślipek & Młynarczuk, 2013), así como en la clasificación y segmentación de minerales en rocas (Saxena et al., 2021; Srinivas Amiripalli et al., 2021). Específicamente, cuando el objetivo de reconocimiento es una imagen de roca de un solo tipo, la probabilidad de reconocimiento es mayor que cuando la imagen presenta varios tipos de rocas (Liu et al., 2020). Las herramientas de aprendizaje automático facilitan el trabajo de geología, permitiendo que este tipo de estudios petrográficos se puedan automatizar para estudiar rocas y minerales, además de ayudar a comprender las estructuras geológicas subyacentes y predecir la fragmentación de rocas y sedimentos en una determinada cuenca aluvial, entre otros (Patro et al., 2022).

Entre los modelos existentes de CNN se encuentra el algoritmo You Only Look Once (YOLO), desarrollado por Ultralytics (Jocher et al., 2023). Este algoritmo se ha destacado por sus mejoras en la precisión, velocidad y capacidad de detección de objetos pequeños en sus diferentes versiones. Distintas versiones se han desarrollado, realizando ajustes en la arquitectura y el entrenamiento para lograr

resultados óptimos en diversos escenarios. Estos han sido tomados como puntos de inicio para el reentrenamiento de modelos para la detección de rocas en labores mineras (Rodríguez-Guillén et al., 2024), identificación de minerales en secciones delgadas (Pratama et al., 2023), detección de deslizamientos a partir de imágenes de sensores remotos (Han et al., 2023) e incluso detección de basura a partir de videos (Carolis et al., 2020).

El objetivo del presente trabajo es la creación de modelos reentrenados basados en YOLOv8 para facilitar la automatización en la identificación y clasificación de diferentes tipos de rocas. La idea es determinar variables de fácil control en la etapa de entrenamiento para así evaluar su utilidad y potencial aplicación a diversos conjuntos de datos que abarcan diferentes categorías de imágenes. Consideramos que el objetivo se logra mediante el uso de YOLO, ya que este algoritmo se destaca por su capacidad de procesamiento rápido y su habilidad para adaptarse fácilmente a nuevas tareas, sin comprometer la calidad (Redmon et al., 2015). La estructura simple de YOLOv8 facilita el proceso de entrenamiento, permitiéndole adaptarse eficazmente a una variedad de conjuntos de datos y aplicaciones, respaldado por una comunidad robusta y un desarrollo continuo. Estos modelos servirán como base para futuros proyectos geocientíficos más específicos y permitirán a instituciones académicas tener un punto de partida para entrenar modelos con sus propias muestras de roca.

## 2. Materiales y métodos

Para el reentrenamiento de los modelos YOLOv8 se recolectaron imágenes de muestras de rocas procedentes de repositorios en línea y del catálogo de muestras de rocas de la Escuela de Geología de la Universidad Industrial de Santander (UIS). Cada imagen fue verificada para asegurar que correspondiera a una roca específica y se clasificó como ígnea, sedimentaria o metamórfica para su posterior etiquetado. El proceso de entrenamiento se realizó variando los bancos de imágenes para generar diversos modelos, los cuales fueron

reentrenados de manera progresiva. Para validar y verificar la eficacia de los modelos, se utilizó un catálogo limitado de muestras de roca de la Escuela de Geología, para de esta manera evaluar las ventajas de cada uno de ellos.

### 2.1. Datasets

Se recolectaron más de tres mil imágenes de rocas como etapa previa al entrenamiento, consultadas de los catálogos disponibles en Kaggle (Tabla 1). Cada imagen fue revisada cuidadosamente para verificar que representara muestras de roca y que su calidad permitiera visualmente una clasificación general. Este proceso de verificación de la calidad de las imágenes de Kaggle fue fundamental dado que muchas presentaban baja resolución. Se priorizaron imágenes con una resolución mínima de 640x640 píxeles, adecuada para la capa de entrada de los modelos YOLO.

**Tabla 1.** Listado de los catálogos de imágenes consultados en Kaggle.

Nombre	Enlace Kaggle
Rock Classification Dataset	<a href="https://www.kaggle.com/datasets/salmaneunus/rock-classification?resource=download">https://www.kaggle.com/datasets/salmaneunus/rock-classification?resource=download</a>
Rock Images	<a href="https://www.kaggle.com/datasets/neelgajare/rocks-dataset?select=Rocks">https://www.kaggle.com/datasets/neelgajare/rocks-dataset?select=Rocks</a>
Igneous metamorphic sedimentary rocks and minerals	<a href="https://www.kaggle.com/datasets/mahmoudalforawi/igneous-metamorphic-sedimentary-rocks-and-minerals?select=Rock_Dataset">https://www.kaggle.com/datasets/mahmoudalforawi/igneous-metamorphic-sedimentary-rocks-and-minerals?select=Rock_Dataset</a>

Además, con el acceso a una parte del catálogo básico de muestras de la Escuela de Geología, se seleccionaron 48 muestras de rocas ígneas, 34 muestras de rocas sedimentarias y 15 muestras de rocas metamórficas. Cada muestra fue fotografiada individualmente con un celular Samsung A13, utilizando una resolución de 8160x6120 píxeles para la captura de imágenes, lo que facilitó una identificación general del tipo de roca correspondiente entre los tres tipos mencionados.

## 2.2. Etiquetado

Para el etiquetado de las muestras en la categoría de roca ígnea, roca metamórfica o roca sedimentaria se utilizó la herramienta de anotación de imágenes gráfica Labelme (Wada et al., 2021). Cada muestra de roca fue delimitada mediante un polígono irregular y se asignó la etiqueta correspondiente al cerrar el polígono. Igualmente, cada imagen fue etiquetada manualmente. Además, como parte de este proceso, se realizó un aumento de datos manual colocando los tres tipos de rocas en una única imagen, únicamente con muestras de la UIS (Figura 1).

## 2.3. Entrenamiento

La etapa de entrenamiento o reentrenamiento de los modelos fue realizada en el entorno virtual de Google Colaboratory, comúnmente denominado “Google Colab” o simplemente “Colab”, específicamente la versión PRO. Los recursos utilizados corresponden al entorno L4, conformado por 53.0 GB de RAM, 22.5 GB de GPU y 78.2 GB de Disco.

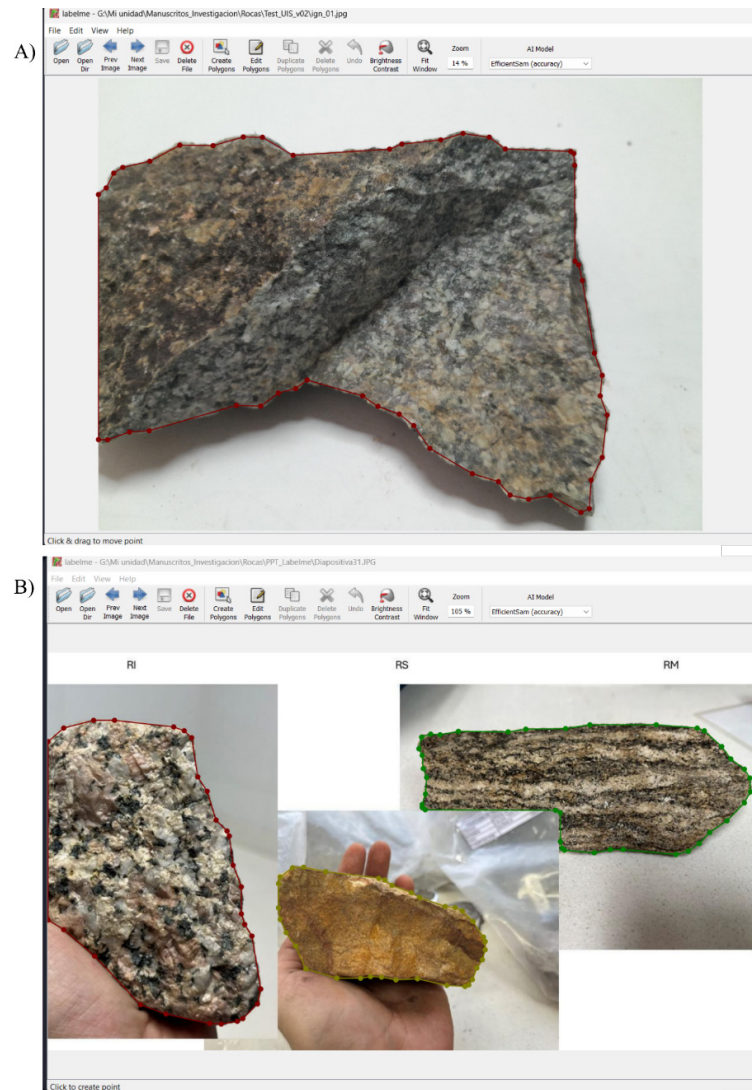
Para esta tarea, dos modelos de segmentación de YOLOv8 de la librería de Ultralytics (Jocher et al., 2023) fueron seleccionados: (i) YOLOv8n-seg conformada por 3,4 millones de parámetros y (ii) YOLOv8x-seg con 71,8 millones de parámetros. El primer modelo corresponde al más simple al tener menor cantidad de parámetros. La librería de Ultralytics permite realizar modificaciones a las variables de entrenamiento, como lo son redimensionar las imágenes de entrada, el tamaño del batch o grupo de entrenamiento, modificar las cantidades de épocas de entrenamiento, entre otras. También permite generar un aumento de datos adicional (Figura 2).

Los dos modelos de segmentación seleccionados fueron entrenados en diferentes etapas, en las cuales se realizaron cambios en las variables de entrenamiento y el dataset utilizado. El modelo obtenido en cada etapa fue el punto de partida del próximo reentrenamiento, por lo que en cada etapa se obtuvo un modelo (Tabla 2). La nomenclatura utilizada para distinguir cada modelo e identificar cada etapa de entrenamiento sigue el formato: “rocas\_” indica que se trata de un modelo para la clasificación de rocas. Las letras “k” o “u” indican el orden, de izquierda a derecha, en el que

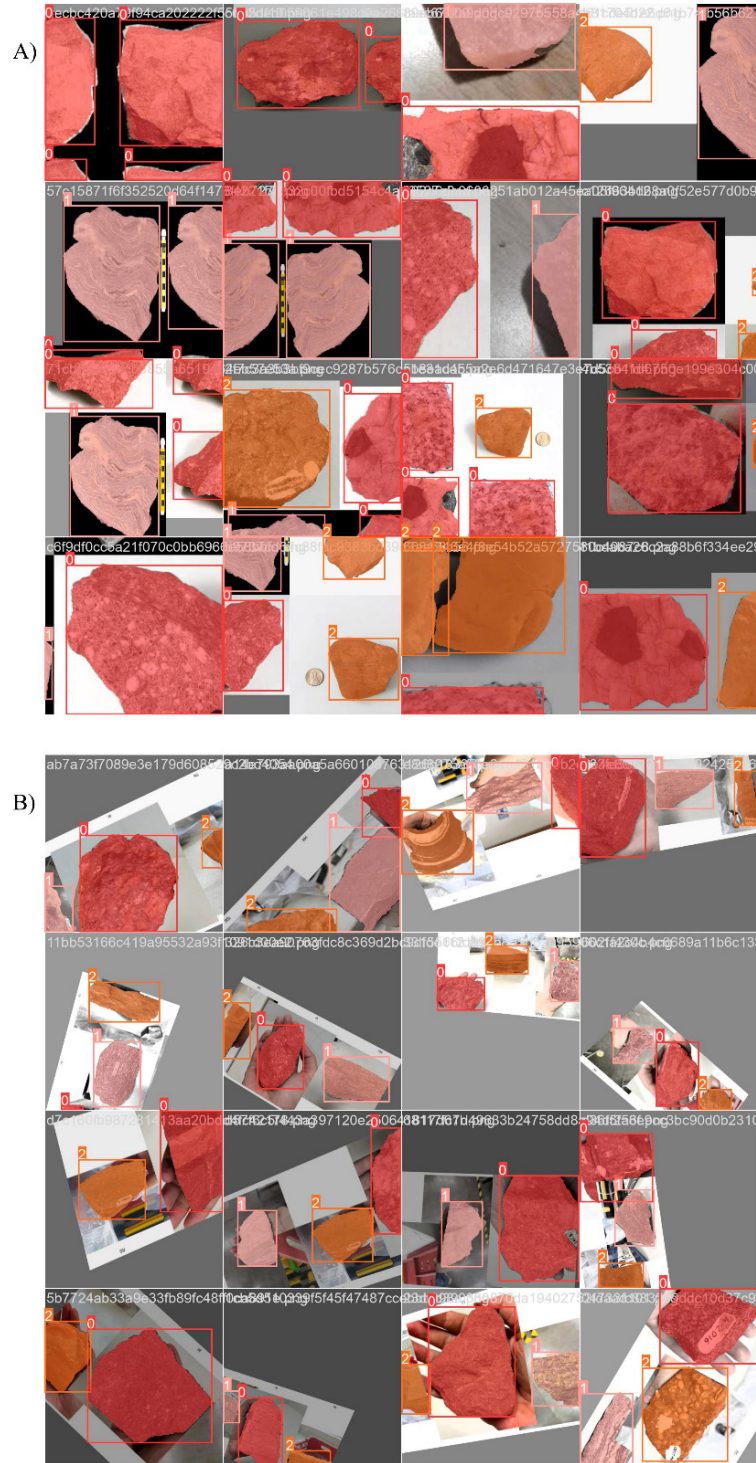


se entrenó con el catálogo respectivo, siendo “k” para Kaggle y “u” para UIS. Finalmente, se añade “-n” para indicar que el modelo base corresponde a YOLOv8n-seg, y “-x” para el modelo base YOLOv8x-seg.

**Figura 1.** Ejemplo de etiquetado utilizando Labelme (Wada et al., 2021) para A) una única muestra de roca y B) múltiples muestras de rocas mediante un aumento de datos manual. Estas muestras hacen parte del catálogo de la UIS.



**Figura 2.** Ejemplo de las imágenes de rocas ígneas (color rojo), sedimentarias (color naranja) y metamórficas (color rosado) en la etapa del entrenamiento de los modelos de YOLOv8, con A) las variables por defecto y B) realizando modificaciones para el aumento de datos.



**Tabla 2.** Características de las etapas de entrenamiento de cada modelo. El signo positivo en las épocas indica que son adicionales a las del modelo obtenido previo.

	Etapa	Modelo	Épocas	Dataset	Aumento de datos
YO- LOv8n-seg	1	rocas_k-n	+100	Kaggle	No
	2	rocas_kk-n	+60	Kaggle	Si
	3	rocas_kku-n	+103	UIS	Si
	4	rocas_kkuu-n	+15	UIS	No
	5	rocas_kkuuk-n	+15	Kaggle	No
	6	rocas_kkuuku-n	+5	UIS	No
YOLO- v8x-seg	1	rocas_k-x	+160	Kaggle	Si
	2	rocas_ku-x	+103	UIS	Si
	3	rocas_kuu-x	+15	UIS	No
	4	rocas_kuuk-x	+15	Kaggle	No
	5	rocas_kuuku-x	+5	UIS	No

#### 2.4. Validación

Para la validación de los modelos generados, con miras a futuros proyectos geocientíficos, se utilizó la matriz de confusión. Esta evaluación se llevó a cabo con nuevas muestras de la UIS que no fueron incluidas durante el entrenamiento inicial. En total, se etiquetaron 138 nuevas muestras para la validación, distribuidas en 29 rocas ígneas, 64 rocas metamórficas y 45 rocas sedimentarias. Debido a las diferencias en la cantidad de cada tipo de roca para la validación, se normalizó la matriz de confusión en función de las columnas o etiquetas reales, asegurando que la suma de cada columna fuera igual a 1. Esto permitió calcular el porcentaje de aciertos o etiquetados correctos para cada tipo de roca, de manera proporcional.

### 3. Resultados y discusión

Se tomó la decisión de realizar diferentes etapas de entrenamiento, variando los conjuntos de datos y aumentando su cantidad, debido a la variabilidad en la calidad y el tamaño de las imágenes del catálogo de Kaggle y de la UIS. Durante la validación de estas imágenes, se

identificó que no todas tenían la misma calidad en el catálogo de Kaggle, pero se contaba con una mayor cantidad de imágenes, por lo que se inició el entrenamiento con este catálogo para que el modelo pudiera generalizar en la identificación de los tres tipos de rocas. Posteriormente, se realizó el entrenamiento con imágenes del catálogo de la UIS, donde todas las imágenes eran consistentes en condiciones y calidad. Debido a la diferencia en la cantidad de imágenes entre los dos catálogos, el inicio del entrenamiento con el repositorio de Kaggle consideró más de 100 épocas para asegurar una mejor generalización del modelo. En las últimas etapas, se volvió a entrenar con el catálogo de Kaggle para evitar un posible sobreajuste únicamente a las muestras de la UIS.

Un total de once modelos reentrenados fueron obtenidos de los modelos de YOLOv8, seis utilizando YOLOv8n (Tabla 3) y cinco utilizando YOLOv8x (Tabla 4). Aunque no se realizó un control estricto de los tiempos de entrenamiento, se observó que los tiempos para el entrenamiento de los modelos basados en YOLOv8n fueron relativamente menores en comparación con los modelos basados en YOLOv8x. Esto es coherente considerando la cantidad de parámetros de cada modelo. Una mayor cantidad de parámetros en los modelos implica más cálculos para ser realizados y ajustados, lo que también afecta la etapa de despliegue del modelo en cualquier proceso de automatización. Un modelo más complejo o con una mayor cantidad de parámetros requiere más espacio en disco y mayores recursos para ejecutarse de manera adecuada. Esto debe tenerse en cuenta al elegir el modelo para aplicaciones prácticas, donde la eficiencia y el uso de recursos son cruciales.

Durante la etapa de validación de los modelos con las muestras no vistas de la UIS inicialmente, solo uno de los modelos, denominado “rocas\_kkuuku-n”, no logró predecir el tipo de roca. Al verificar los resultados proporcionados por la librería de Ultralytics para cada uno de los entrenamientos, se evidenció que este modelo tampoco realizó predicciones de las muestras durante su entrenamiento. Esto podría sugerir que hubo un problema inherente en el proceso de entrenamiento de dicho modelo, como un conjunto de imágenes insuficiente o una configuración incorrecta de los hiperparámetros. Sin embargo, es importante destacar que el modelo “rocas\_kuuku-x”

también fue entrenado con el mismo conjunto de imágenes e hiperparámetros, y si logró realizar predicciones exitosas. Esto sugiere que el problema no radica únicamente en la cantidad de datos o en la configuración de los hiperparámetros, sino posiblemente en el ajuste de los pesos del modelo. Es posible que, durante el entrenamiento, el modelo haya convergido a lo que se denomina un mínimo local en lugar de alcanzar el mínimo global (Kawaguchi et al., 2019), lo que afectaría su capacidad para generalizar correctamente las predicciones. La posibilidad se tiene en cuenta porque ésta fue la última etapa del entrenamiento con solo cinco épocas, lo que tal vez no fue suficiente para que el modelo escapara del mínimo local y explorara soluciones más óptimas en el espacio de parámetros. Aun así, considerando que el modelo ya ha pasado por entrenamientos previos, donde logró generalizar la identificación de las rocas, no está claro por qué ocurrió esta discrepancia en la etapa final de entrenamiento. No se puede proporcionar una justificación definitiva y segura para este comportamiento del modelo. Muchos investigadores todavía consideran las CNN como un modelo de extracción de características de caja negra (Xin & Wang, 2019). Estudios recientes se han enfocado en desarrollar métodos para comprender qué están aprendiendo las CNN en cada una de sus capas, permitiendo comprender la toma de decisiones de estos modelos (LI et al., 2022; Szandata, 2023).

En el proceso de validación, al evaluar los valores en las diagonales de las respectivas matrices de confusión, se evidencia que diez de los once modelos obtuvieron un acierto superior al 50% en la clasificación de las rocas ígneas. Seis modelos lograron más del 50% de aciertos en la clasificación de las rocas sedimentarias y solo tres modelos alcanzaron más del 50% en la clasificación de las rocas metamórficas. Los modelos con mayores valores en la diagonal de las matrices de confusión, y que pueden ser considerados como los que mejor generalizan en la clasificación de las rocas, fueron “rocas\_kkuuk-n” (Figura 3) y “rocas\_kuuku-x” (Figura 4).

**Tabla 2.** Características de las etapas de entrenamiento de cada modelo. El signo positivo en las épocas indica que son adicionales a las del modelo obtenido previo.

	Etapa	Modelo	Épocas	Dataset	Aumento de datos
YO- LOv8n-seg	1	rocas_k-n	+100	Kaggle	No
	2	rocas_kk-n	+60	Kaggle	Si
	3	rocas_kku-n	+103	UIS	Si
	4	rocas_kkuu-n	+15	UIS	No
	5	rocas_kkuuk-n	+15	Kaggle	No
	6	rocas_kkuuku-n	+5	UIS	No
YOLO- v8x-seg	1	rocas_k-x	+160	Kaggle	Si
	2	rocas_ku-x	+103	UIS	Si
	3	rocas_kuu-x	+15	UIS	No
	4	rocas_kuuk-x	+15	Kaggle	No
	5	rocas_kuuku-x	+5	UIS	No

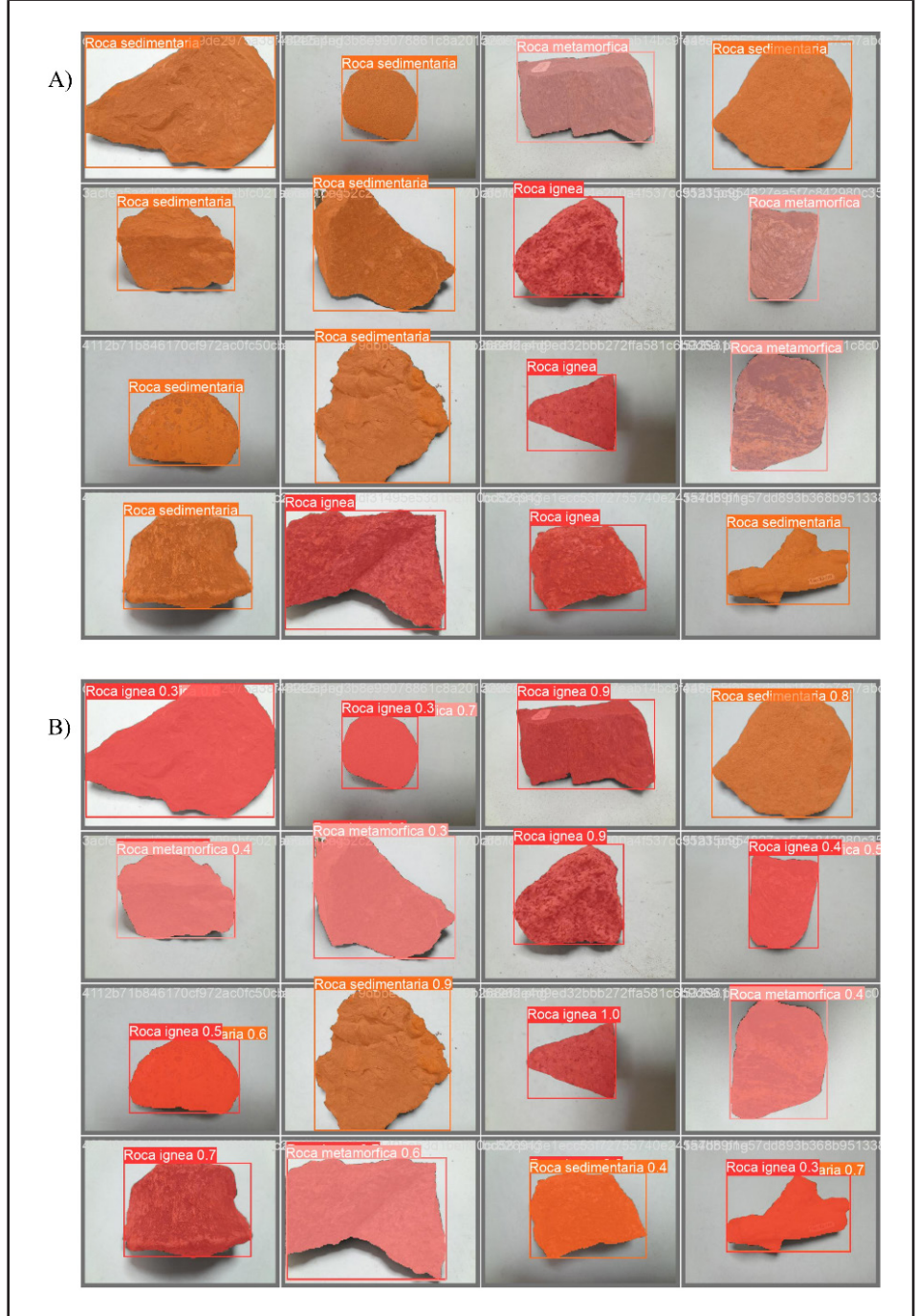
**Tabla 3.** Matriz de Confusión Normalizada, con las etiquetas verdaderas para seis modelos entrenados, usando YOLOv8n. Celdas en color celeste valores normalizados entre 0,25 y 0,50, celdas en verde valores normalizados mayores o iguales a 0,50.

Modelo		RI_true	RM_true	RS_true
rocas_k-n	RI_pred	0,62	0,48	0,38
	RM_pred	0,07	0,19	0,02
	RS_pred	0,31	0,33	0,60
rocas_kk-n	RI_pred	0,52	0,30	0,22
	RM_pred	0,17	0,25	0,07
	RS_pred	0,31	0,45	0,71
rocas_kku-n	RI_pred	0,83	0,47	0,62
	RM_pred	0,10	0,53	0,22
	RS_pred	0,07	0,00	0,16
rocas_kkuu-n	RI_pred	0,66	0,50	0,31
	RM_pred	0,10	0,42	0,27
	RS_pred	0,03	0,00	0,09
rocas_kkuuk-n	RI_pred	0,66	0,45	0,27
	RM_pred	0,24	0,34	0,09
	RS_pred	0,10	0,20	0,64
rocas_kkuuku-n	RI_pred	0,00	0,00	0,00
	RM_pred	0,00	0,00	0,00
	RS_pred	0,00	0,00	0,00

**Tabla 4.** Matriz de Confusión Normalizada, con las etiquetas verdaderas para cinco modelos entrenados, usando YOLOv8x. Celdas en color celeste valores normalizados entre 0,25 y 0,50, celdas en verde valores normalizados mayores o iguales a 0,50.

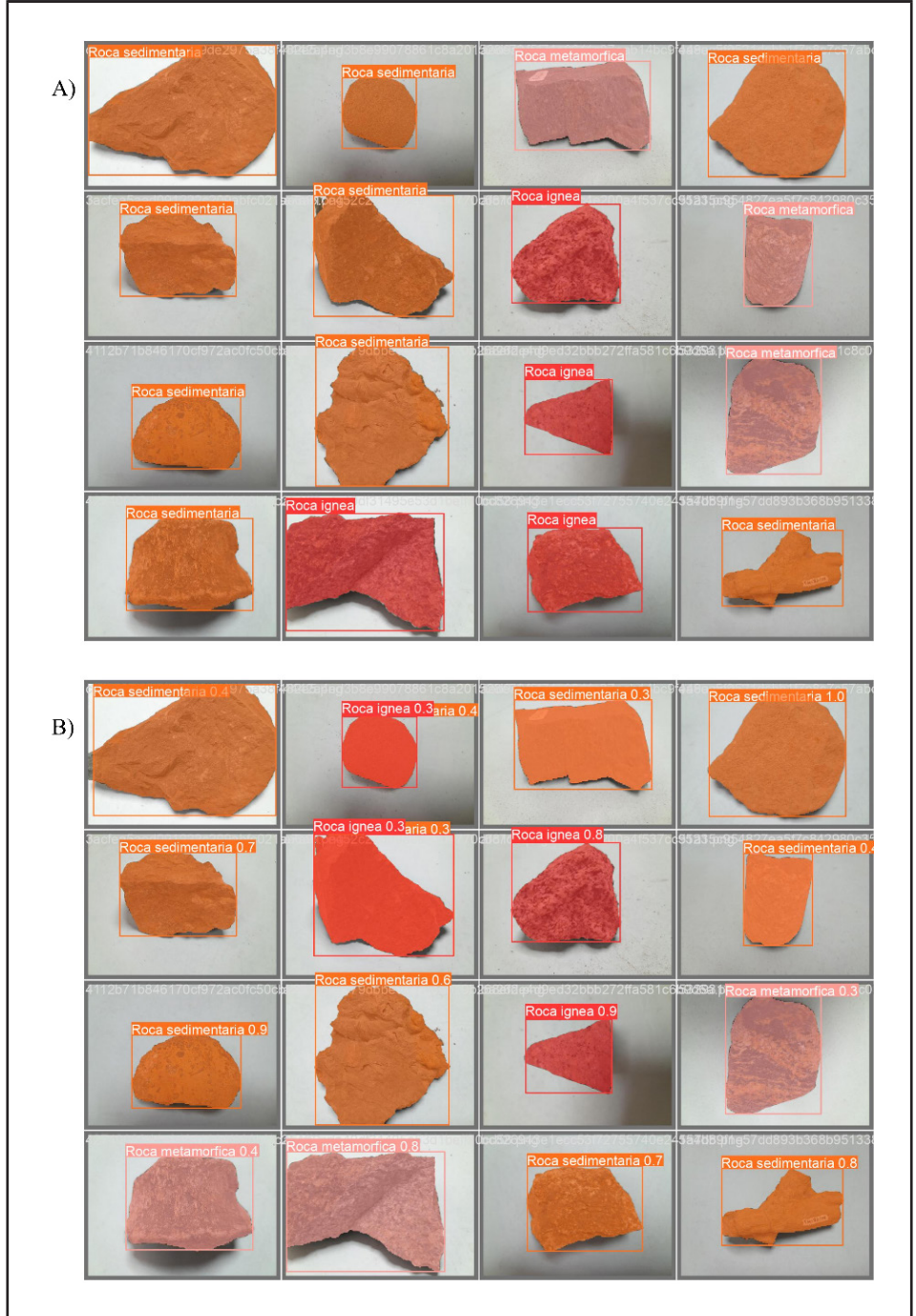
Modelo		RI_true	RM_true	RS_true
rocas_k-x	RI_pred	0,55	0,28	0,09
	RM_pred	0,07	0,14	0,02
	RS_pred	0,38	0,58	0,89
rocas_ku-x	RI_pred	0,86	0,41	0,49
	RM_pred	0,14	0,56	0,38
	RS_pred	0,00	0,03	0,13
rocas_kuu-x	RI_pred	0,72	0,36	0,47
	RM_pred	0,24	0,61	0,27
	RS_pred	0,03	0,03	0,24
rocas_kuuk-x	RI_pred	0,55	0,14	0,04
	RM_pred	0,07	0,13	0,02
	RS_pred	0,38	0,73	0,93
rocas_kuuku-x	RI_pred	0,62	0,22	0,18
	RM_pred	0,10	0,38	0,04
	RS_pred	0,24	0,41	0,76

**Figura 3.** Validación final realizada al modelo "rocas\_kkuuk-n" utilizando muestras de la Escuela de Geología de la UIS. En la parte superior, A) las etiquetas originales de roca ígnea (color rojo), sedimentaria (color naranja) y metamórfica (color rosado). En la parte inferior, B) la predicción realizada por el modelo. La muestra en la fila 3, columna 1, es un ejemplo de una muestra con una predicción o etiqueta doble.





**Figura 4.** Validación final realizada al modelo "rocas\_kuuku-x" utilizando muestras de la Escuela de Geología de la UIS. En la parte superior, A) las etiquetas originales de roca ígnea (color rojo), sedimentaria (color naranja) y metamórfica (color rosado). En la parte inferior, B) la predicción realizada por el modelo. La muestra en la fila 1, columna 2, es un ejemplo de una muestra con una predicción o etiqueta doble.



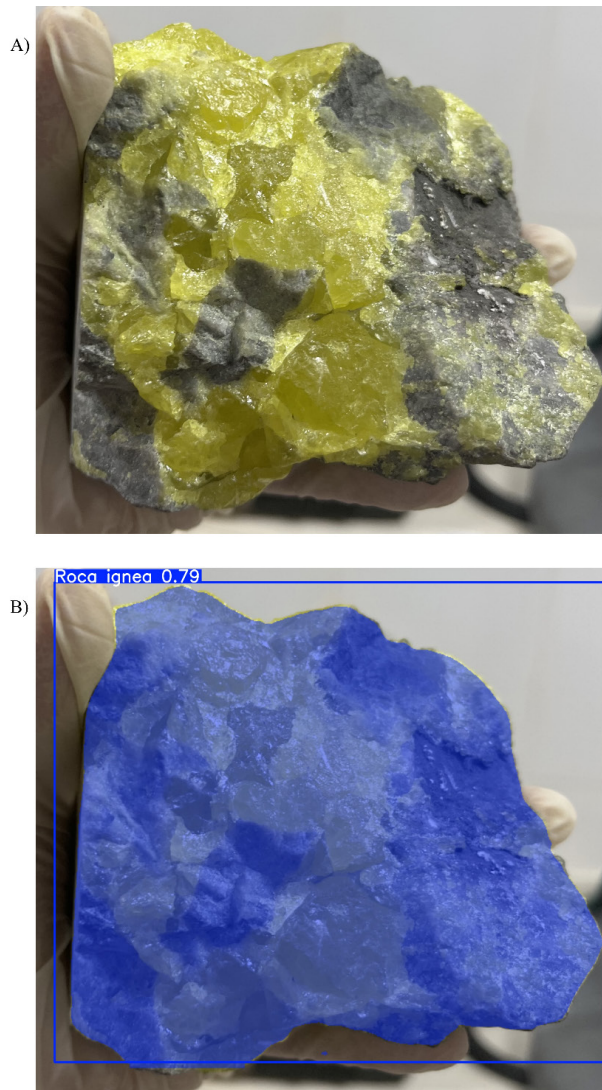
Como se planteó para este trabajo, se pueden seleccionar modelos específicos como punto de partida para futuros proyectos. Por ejemplo, si se desea entrenar modelos que consideren solo rocas ígneas, se podrían utilizar los modelos “rocas\_kku-n” y “rocas\_ku-x”, los cuales lograron una precisión superior al 80% en la etapa de validación. De manera similar, el modelo que obtuvo el mayor acierto en la identificación de rocas metamórficas fue “rocas\_kuu-x”, con un 61%. Para la identificación de rocas sedimentarias, el modelo “rocas\_kuuk-x” logró un porcentaje de acierto del 93%. Dependiendo de los objetivos de futuros proyectos, se podría seleccionar un modelo específico con mayores puntajes en una de las categorías de rocas o una combinación balanceada de los mismos.

Los modelos de YOLOv8 están diseñados para generar valores entre 0,0 y 1,0 para cada categoría, lo que representan la probabilidad o porcentaje de confiabilidad de la etiqueta asignada. En este caso, se puede obtener la probabilidad de que cada muestra pertenezca a uno de los tres tipos de roca, siendo la etiqueta con el valor más alto, la considerada como la predicción del modelo. Sin embargo, también se puede permitir al modelo que presente más de una etiqueta si el porcentaje supera un umbral específico. Durante la etapa de validación se observó que en algunas muestras el modelo devolvía más de una etiqueta (Figura 4), por lo que, en procesos de automatización, se puede considerar esta posibilidad para una posterior verificación de resultados por control humano, lo que puede mejorar la precisión y confiabilidad de las predicciones del modelo.

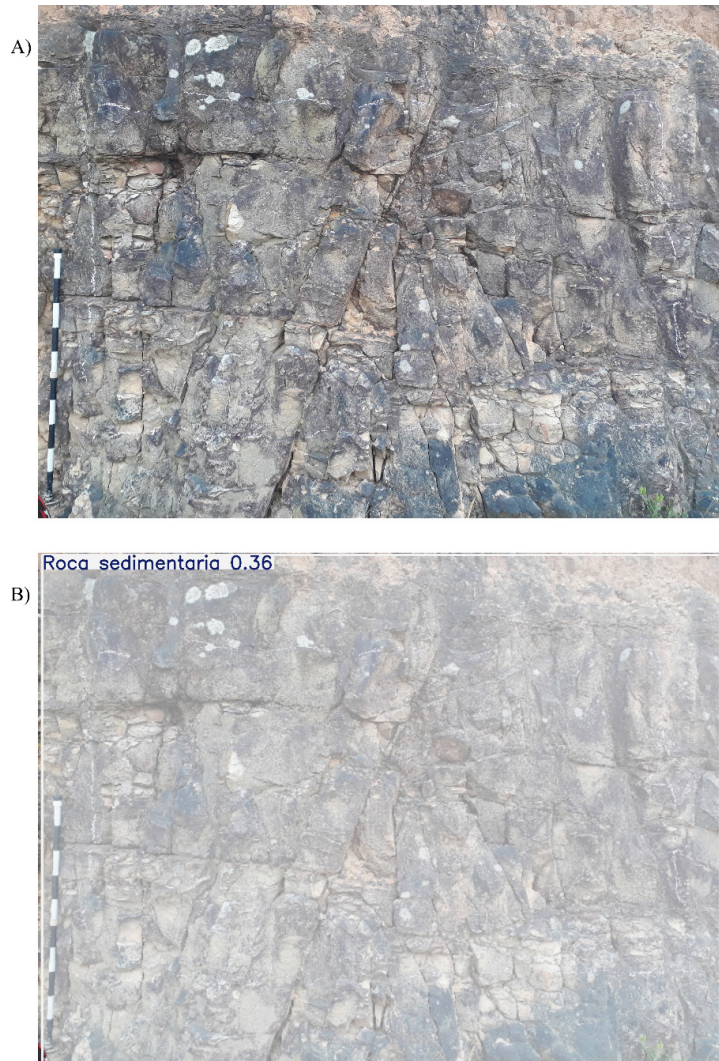
Los modelos de CNN, implementados para el reconocimiento litológico, pueden identificar el tipo de roca con imágenes incompletas o complejas, lo que demuestra la robustez y capacidad de generalización de las CNN (Liu et al., 2020). Entre las consideraciones de los modelos está la posibilidad de que un usuario de un modelo de CNN ingrese una imagen que no corresponda directamente al objetivo del modelo, como podría ser una imagen de un mineral (Figura 5). En tal caso, el modelo podría asignar una etiqueta en las tres categorías de roca, lo cual no sería estrictamente correcto. Por esta razón, es crucial ingresar únicamente las imágenes correspondientes o incluir nuevas categorías en el modelo. Para

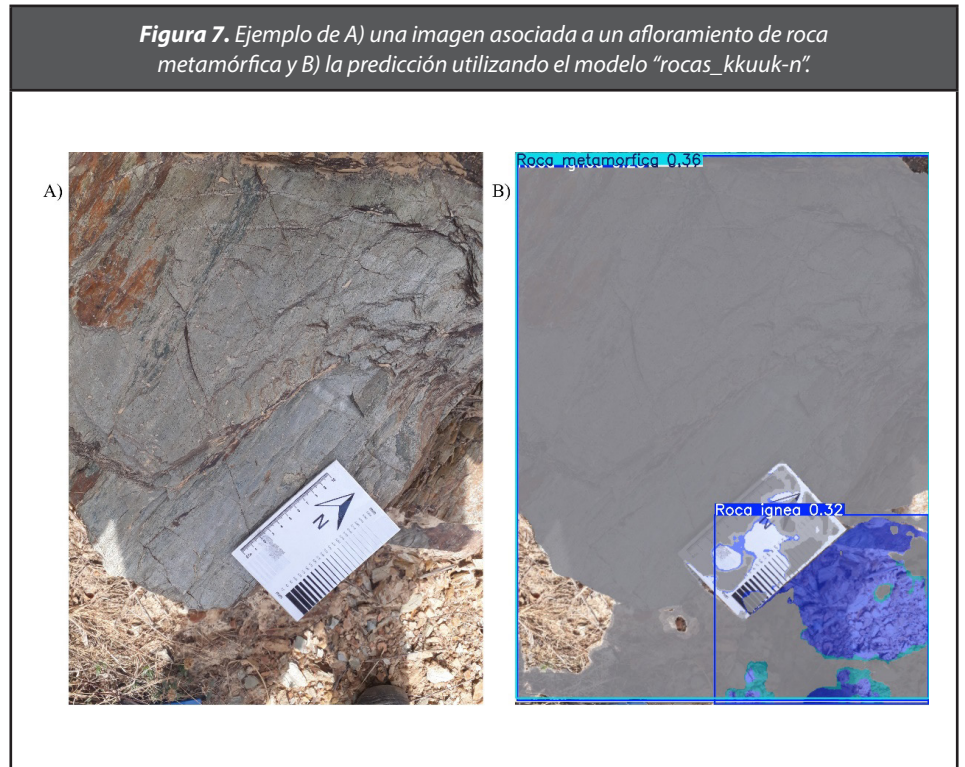
abordar esta situación, lo más recomendable es reentrenar un modelo para que considere minerales en su clasificación. Siguiendo esta misma idea, los modelos entrenados pueden recibir imágenes de un afloramiento (Figura 6, Figura 7) y dependiendo de la calidad de la imagen, podría ser posible clasificar el afloramiento en un tipo de roca específico, pero es más conveniente que los modelos sean entrenados para que además analicen específicamente las imágenes de afloramientos.

**Figura 5.** Ejemplo de A) una imagen correspondiente a un mineral de azufre y B) la respectiva predicción utilizando el modelo "rocas\_kkuuk-n".



**Figura 6.** Ejemplo de A) una imagen asociada a un afloramiento de roca sedimentaria y B) la predicción utilizando el modelo "rocas\_kkuuk-n".





Para el despliegue o puesta en producción de un modelo, es necesario llevarlo a diferentes plataformas y cambiar su formato de salida según las necesidades del sistema. Otra de las bondades de utilizar estos modelos es que Ultralytics (Jocher et al., 2023) ofrece diferentes formatos para exportar los modelos basados en YOLOv8, lo que permite realizar modificaciones adicionales en la estructura del modelo, como ajustar el número de capas incluyendo o retirando las mismas en librerías externas, por ejemplo, TensorFlow (Abadi et al., 2015). Esta opción brinda flexibilidad para adaptar, de ser requerido, el modelo a diferentes entornos y requerimientos específicos de la aplicación en producción. Adicionalmente, a la fecha YOLOv8 redimensiona las imágenes a 640x640 píxeles, lo que permite que entradas de mayor tamaño se puedan realizar en librerías externas.

El beneficio del Transfer Learning radica en la capacidad de tomar un modelo previamente entrenado y luego reentrenarlo, lo que puede reducir significativamente las épocas o el tiempo necesario para entrenar nuevos modelos. Esto es posible porque muchas características de la imagen ya están extraídas en el modelo previo, permitiendo que, con mayor eficiencia, el nuevo modelo se

enfoque en aprender las específicas del nuevo conjunto de datos. Esto proporciona una precisión de clasificación razonable incluso con un conjunto de reentrenamiento de baja cantidad (Hussain et al., 2019).

Una de las modificaciones posibles es eliminar la capa final del modelo previamente entrenado, lo que permite adaptar el modelo de manera más efectiva a la tarea específica de clasificación de rocas. Para datasets de baja cantidad, como las muestras de la UIS utilizadas en este trabajo, el aumento de datos manual es de gran utilidad. Adicionalmente, se pueden incluir otras modificaciones controladas para incrementar la variedad del dataset, por ejemplo, probar estrategias como: (i) rotar las muestras o cambiar la perspectiva desde la cual se toma la foto, lo que resulta en imágenes de entrenamiento diferentes; (ii) mezclar diferentes muestras en una misma imagen, lo que puede ayudar a entrenar un modelo para que generalice adecuadamente; (iii) variar el fondo de la muestra, es una modificación que se puede realizar de manera manual y controlada. Estos cambios ayudan a incrementar el dataset, mejorar la robustez y capacidad de generalización del modelo.

La librería de Ultralytics ofrece varias técnicas de aumento de datos, como la rotación de las imágenes, recortes, mezcla de imágenes, eliminación de píxeles, entre otras. Es importante realizar estas modificaciones de manera controlada para no comprometer la calidad de la imagen hasta el punto de que ni siquiera de parte de profesionales de la geología pueda ser clasificada correctamente.

Todas las consideraciones mencionadas y previas al entrenamiento son importantes para integrar el modelo en sistemas que requieren un tamaño específico de imagen de entrada. Esto asegura una compatibilidad adecuada y un rendimiento óptimo del modelo en producción, garantizando que las imágenes procesadas cumplan con los requisitos necesarios para una clasificación precisa y eficiente.

El uso de CNN puede resultar útil en los sistemas de registro de clases, implementaciones del sistema en un servidor web o Internet de las cosas (IoT) (Hussain et al., 2019). Todavía hay margen para mejorar los resultados para proyectos más especializados, mediante la implementación de modelos con un mayor número de capas, un

aumento en las épocas de entrenamiento y la realización de pruebas con conjuntos de datos más extensos y específicos. Uno de los proyectos más cercanos es la utilización y alimentación continua de un modelo con las muestras que ingresan a un catálogo de una institución provenientes de las diferentes salidas de campo, incluyendo nuevas etiquetas para los respectivos tipos de clasificación. Este proyecto permitirá no solo mejorar la precisión del modelo, sino mantenerlo actualizado con nuevas variaciones y tipos de rocas, asegurando así su relevancia y aplicabilidad en investigaciones geocientíficas futuras. Es importante recordar que las tareas de entrenamiento de modelos de CNN requieren tiempo y recursos significativos, especialmente en la etapa de etiquetado y verificación de las imágenes utilizadas para el entrenamiento del modelo.

#### 4. Conclusiones

La implementación de diferentes etapas de entrenamiento, que incluyen imágenes de rocas de Kaggle y del catálogo de la UIS, permite variar la diversidad de condiciones y calidad de las imágenes. Iniciar con el catálogo de Kaggle proporciona una base generalizada, mientras que el uso del catálogo de la UIS tiene como objetivo refinar el modelo y adaptarlo a muestras de roca más detalladas. Esto resalta la importancia de adaptar las estrategias de entrenamiento a las características específicas de los datos disponibles, incluyendo la posibilidad de aumentar los datos de forma manual o automática utilizando las herramientas disponibles en las bibliotecas de entrenamiento.

El reentrenamiento facilita la generación de nuevos modelos al requerir menos iteraciones. En este trabajo se lograron generar once modelos, dos de los cuales fueron “rocas\_kkuuk-n” y “rocas\_kuuku-x”, con valores mayores al 30% en la diagonal de las matrices de confusión que pueden ser considerados como los que generalizan mejor en la clasificación de las rocas. Aun así, los modelos “rocas\_kku-n” y “rocas\_ku-x” lograron una precisión superior al 80% en la etapa de validación para la identificación de rocas ígneas. Para la identificación de rocas sedimentarias, el modelo “rocas\_kuuk-x”

logró un porcentaje de acierto del 93%. El modelo “rocas\_kuu-x” obtuvo el mayor acierto en la identificación de rocas metamórficas con un 61%. Estos modelos pueden ser la base para desarrollar otros más complejos o específicos, lo cual se puede lograr agregando nuevas capas o realizando otras modificaciones a la estructura de la CNN, adaptando así los modelos a las necesidades particulares de cada proyecto. En el siguiente enlace de Drive se disponen los once modelos generados: <https://drive.google.com/drive/folders/1u9ZhHxB7W6h5A4b4G3sV5qoR2HdmLrVd?usp=sharing>. Para conocer parte del código se puede consultar el repositorio de Github: [https://github.com/sergioGarcia91/YOLO\\_Rocas](https://github.com/sergioGarcia91/YOLO_Rocas).

De los once modelos generados, el único que no logró predecir correctamente en el entrenamiento ni en la validación fue “rocas\_kkuuku-n”. Esto lleva a subrayar la importancia de mantener un control iterativo y almacenar diferentes versiones del modelo para abordar discrepancias inesperadas en el desempeño. Aunque los modelos no fueron entrenados utilizando imágenes de afloramientos, en algunos casos fue posible estimar el tipo de roca al que pertenecen los afloramientos. Así se resalta la capacidad de estos modelos para extraer características básicas. Con entrenamientos adicionales, utilizando catálogos más específicos, se puede mejorar significativamente el rendimiento de los modelos, para que requieran una menor cantidad de épocas, sobre todo si se compara con el entrenamiento desde cero con un modelo no especializado en geociencias.

La validación realizada mediante la matriz de confusión normalizada permitió identificar de manera simple los modelos con mayores fortalezas en la identificación de un tipo de rocas. El proceso de ajuste, validación y eventual despliegue de estos modelos requiere consideraciones cuidadosas en términos de recursos computacionales, eficiencia y adaptabilidad de las plataformas de producción. Los modelos más complejos, con una mayor cantidad de parámetros, implican un mayor poder de cómputo y tiempos de entrenamiento más largos, pero que pueden estar disponibles en las instituciones educativas para estos fines.



## 5. Contribución de los autores

**Sergio García-Arias:** Investigación, recolección y preparación de imágenes, metodología, análisis formal, redacción, revisión y edición.

**Francisco Velandia:** Metodología, análisis formal, redacción, revisión y edición.

## 6. Agradecimientos

Se agradece a la Escuela de Geología y al grupo de Geología Básica y Aplicada (GIGBA), Universidad Industrial de Santander por brindar acceso a parte del catálogo de rocas utilizado en el entrenamiento y la validación de los modelos.

## 7. Conflictos de interés

Los autores declaran que no tienen intereses financieros en competencia conocidos ni relaciones personales que pudieran haber parecido influir en el trabajo informado en este documento.

## Referencias

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., S.Corrado, G., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Zenodo. <https://doi.org/10.5281/zenodo.12119782>
- Carolis, B. De, Ladogana, F., & Macchiarulo, N. (2020). YOLO TrashNet: Garbage Detection in Video Streams. *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 1–7. <https://doi.org/10.1109/EAIS48028.2020.9122693>
- Cheng, G., & Guo, W. (2017). Rock Images Classification by Using Deep Convolution Neural Network. *Journal of Physics: Conference Series*, 887, 012089. <https://doi.org/10.1088/1742-6596/887/1/012089>

- Cracknell, M. J. (2014). *Machine Learning for Geological Mapping: Algorithms and Applications* [Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy, University of Tasmania]. <https://doi.org/https://doi.org/10.25959/23236064.v1>
- Dramsch, J. S. (2020). 70 Years of Machine Learning in Geoscience in Review. En *Advances in Geophysics* (pp. 1–55). Elsevier Inc. <https://doi.org/10.1016/bs.agph.2020.08.002>
- García-Arias, S., Velandia, F., Sanabria Gómez, J. D., & Ulloque Ardila, M. T. (2024). Mapas Predictivos con Redes Neuronales a Partir de Propiedades Físicas de Las Rocas: Caso de Estudio en la Mesa de Los Santos (Santander). *Revista EIA*, 21(41). <https://doi.org/10.24050/reia.v21i41.1731>
- Gómez, J., Sánchez, J., Ocampo, A., & Restrepo, J. W. (2013). Aplicación de Redes Neuronales en la Clasificación de Arcillas. *Revista EIA*, 9(17), 183–191. <https://revistas.eia.edu.co/index.php/reveia/article/view/459>
- Goyes-Peñañiel, P., & Hernandez-Rojas, A. (2021). Doble Evaluación de la Susceptibilidad por Movimientos en Masa Basada en Redes Neuronales Artificiales y Pesos de Evidencia. *Boletín de Geología*, 43(1). <https://doi.org/10.18273/revbol.v43n1-2021009>
- Han, Z., Fang, Z., Li, Y., & Fu, B. (2023). A novel Dynahead-Yolo Neural Network for the Detection of Landslides with Variable Proportions Using Remote Sensing Images. *Frontiers in Earth Science*, 10. <https://doi.org/10.3389/feart.2022.1077153>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Hussain, M., Bird, J. J., & Faria, D. R. (2019). *A Study on CNN Transfer Learning for Image Classification* (pp. 191–202). [https://doi.org/10.1007/978-3-319-97982-3\\_16](https://doi.org/10.1007/978-3-319-97982-3_16)
- Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLOv8*. <https://github.com/ultralytics/ultralytics>
- Kawaguchi, K., Huang, J., & Kaelbling, L. P. (2019). Every Local Minimum Value Is the Global Minimum Value of Induced Model in Nonconvex Machine Learning. *Neural Computation*, 31(12), 2293–2323. [https://doi.org/10.1162/neco\\_a\\_01234](https://doi.org/10.1162/neco_a_01234)
- Li, J., Wu, Z., & Shen, Z. (2022). Open the Black Box – Visualising CNN to Understand Its Decisions on Road Network Performance Level. *Promet*, 34(4). <https://doi.org/10.7307/ptt.v34i4.4037>
- Liu, X., Wang, H., Jing, H., Shao, A., & Wang, L. (2020). Research on Intelligent Identification of Rock Types Based on Faster R-CNN Method. *IEEE Access*, 8, 21804–21812. <https://doi.org/10.1109/ACCESS.2020.2968515>
- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge.

- Młynarczuk, M., Górszczyk, A., & Ślipek, B. (2013). The Application of Pattern Recognition in the Automatic Classification of Microscopic Rock Images. *Computers & Geosciences*, 60, 126–133. <https://doi.org/10.1016/j.cageo.2013.07.015>
- Patro, S., Jhariya, D. C., Sahu, M., Dewangan, P., & Dhekne, P. Y. (2022). Igneous Rock Classification Using Convolutional Neural Networks (CNN). *IOP Conference Series: Earth and Environmental Science*, 1032(1), 012045. <https://doi.org/10.1088/1755-1315/1032/1/012045>
- Pratama, B. G., Qodri, M. F., & Sugarbo, O. (2023). Building YoloV4 Models for Identification of Rock Minerals in Thin Section. *IOP Conference Series: Earth and Environmental Science*, 1151(1), 012046. <https://doi.org/10.1088/1755-1315/1151/1/012046>
- Redmon, J., Divvala, S. K., Girshick, R. B., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *CoRR*, *abs/1506.02640*. <http://arxiv.org/abs/1506.02640>
- Rodríguez-Guillen, R., Kern, J., & Urrea, C. (2024). Fast Rock Detection in Visually Contaminated Mining Environments Using Machine Learning and Deep Learning Techniques. *Applied Sciences*, 14(2), 731. <https://doi.org/10.3390/app14020731>
- Rohwer, R., Wynne-Jones, M., & Wysotzki, F. (1994). Neural Networks. En *Machine Learning, Neural and Statistical Classification* (pp. 84–106). Ellis Horwood.
- Rojas, R. (1996). *Neural Networks*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-61068-4>
- Saxena, N., Day-Stirrat, R. J., Hows, A., & Hofmann, R. (2021). Application of Deep Learning for Semantic Segmentation of Sandstone Thin Sections. *Computers & Geosciences*, 152, 104778. <https://doi.org/10.1016/j.cageo.2021.104778>
- Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Noguees, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298. <https://doi.org/10.1109/TMI.2016.2528162>
- Ślipek, B., & Młynarczuk, M. (2013). Application of Pattern Recognition Methods to Automatic Identification of Microscopic Images of Rocks Registered Under Different Polarization and Lighting Conditions. *Geology, Geophysics & Environment*, 39(4), 373. <https://doi.org/10.7494/geol.2013.39.4.373>
- Srinivas Amiripalli, S., Nageshwara Rao, G., Behara, J., Sanjay Krishna, K., & Pavan Venkat Durga Ram, M. (2021). *Mineral Rock Classification Using Convolutional Neural Network*. <https://doi.org/10.3233/APC210235>
- Szandała, T. (2023). Unlocking the Black Box of CNNs: Visualising the Decision-Making Process with PRISM. *Information Sciences*, 642, 119162. <https://doi.org/10.1016/j.ins.2023.119162>

- Tarback, E. J., & Lutgens, F. K. (2005). *Ciencias de la Tierra: una Introducción a la Geología Física* (8a ed.). Pearson Educación S. A.
- Van der Baan, M., & Jutten, C. (2000). Neural Networks in Geophysical Applications. *GEOPHYSICS*, 65(4), 1032–1047. <https://doi.org/10.1190/1.1444797>
- Wada, K., mpitid, Buijs, M., N., Z. Ch., Kubovčík, Bc. M., Myczko, A., latentix, Zhu, L., Yamaguchi, N., Fujii, S., iamgd67, IlyaOvodov, Patel, A., Clauss, C., Kuroiwa, E., Iyengar, R., Shilin, S., Malygina, T., ... Toft, H. (2021). *wkentaro/labelme: v4.6.0*. Zenodo. <https://doi.org/10.5281/zenodo.5711226>
- Xin, M., & Wang, Y. (2019). Research on Image Classification Model Based on Deep Convolution Neural Network. *EURASIP Journal on Image and Video Processing*, 2019(1), 40. <https://doi.org/10.1186/s13640-019-0417-8>