

Identificación taxonómica no destructiva de *ammonites* del Cenomaniano-Coniaciano en Colombia mediante inteligencia artificial y aprendizaje profundo

Javier García-Guerrero¹

Ricardo Antonio Tobón²

Artículo de investigación



Fecha de recepción: 5 de mayo de 2025 ■ **Fecha de aceptación:** 22 de julio de 2025 ■ **Fecha de publicación:** 12 de junio de 2026

✉ Facultad de Ingeniería, Departamento de Ingeniería Civil, Fundación Universitaria Agraria de Colombia, Bogotá, Colombia.
conto.david@uniagraria.edu.co

García-Guerrero, J., & Tobón, R. A. (2026). Identificación taxonómica no destructiva de ammonites del Cenomaniano-Coniaciano en Colombia mediante inteligencia artificial y aprendizaje profundo. *Revista de Investigaciones Uniagraria*, 13(1), 41-54.

Resumen

La clasificación de macrofósiles —especialmente ammonites— suele requerir mucho tiempo y pericia, y se ve limitada por la calidad de preservación, la plasticidad morfológica y la escasez de colecciones digitalizadas. Estas restricciones han impedido aprovechar plenamente, en los modelos bioestratigráficos regionales, unidades como Hondita, Loma Gorda y Frontera. Para cerrar esta brecha, el presente trabajo introduce un marco metodológico innovador para la identificación taxonómica automatizada y no destructiva de *ammonites* del Cenomaniano-Coniaciano en formaciones del Cretácico Superior de Colombia, mediante inteligencia artificial (IA). Con tal propósito, se propone un modelo de clasificación de imágenes basado en una CNN (ResNet-18 con transfer learning) y entrenado con imágenes etiquetadas por especialistas y enriquecidas con metadatos estratigráficos, que arroja resultados prometedores en precisión e interpretabilidad morfológica. El desarrollo se sustenta en una revisión sistemática de aplicaciones recientes de IA en paleontología —segmentación de fósiles, generación sintética de datos y técnicas no invasivas, como micro-CT y fluorescencia estimulada por láser—. Además de clasificar, el marco facilita integrar información geoquímica y estratigráfica para refinar la cronoestratigrafía y reconstruir paleoambientes; también, ofrece una vía escalable para digitalizar y curar colecciones en contextos con recursos limitados. Al incorporar *machine learning* al flujo de trabajo, el estudio contribuye a modernizar la práctica paleontológica e impulsa la inclusión del patrimonio fósil de Colombia en bases geocientíficas globales.

Palabras clave: *ammonites*, aprendizaje profundo, clasificación de fósiles, cretácico colombiano, identificación taxonómica automatizada, transferencia de aprendizaje.

¹ Facultad de Ingeniería, Departamento de Ingeniería Civil, Fundación Universitaria Agraria de Colombia, Bogotá, Colombia.

² Facultad de Ingeniería, Departamento de Ingeniería Civil, Fundación Universitaria Agraria de Colombia, Bogotá, Colombia.

Non-Destructive Taxonomic Identification of Cenomanian-Coniacian Ammonites in Colombia Using Artificial Intelligence and Deep Learning

Abstract

This paper introduces an innovative methodological framework for the automated and non-destructive taxonomic identification of Cenomanian-Coniacian ammonites from Upper Cretaceous formations in Colombia using artificial intelligence. The taxonomic classification of macrofossils—particularly ammonites—is traditionally a time-consuming and expertise-dependent task, often limited by preservation quality, morphological plasticity, and the scarcity of digitized reference collections. These constraints have hindered the full integration of ammonite-bearing units such as Hondita, Loma Gorda, and Frontera into regional biostratigraphic models. To address this gap, we propose an image classification model based on a convolutional neural network (ResNet-18 with transfer learning), trained on expert-labeled fossil images enriched with stratigraphic metadata, which yields promising results in classification accuracy and morphological interpretability. The approach builds on a systematic review of recent AI applications in paleontology, including fossil segmentation, synthetic dataset generation, and non-invasive imaging techniques such as micro-CT and laser-stimulated fluorescence. Beyond classification, the framework enables the integration of geochemical and stratigraphic data to refine chronostratigraphy and reconstruct paleoenvironments, while offering a scalable solution for digitizing and curating fossil collections in resource-constrained settings. By embedding machine learning into the fossil identification workflow, this study contributes to the modernization of paleontological research and promotes the inclusion of Colombia's fossil heritage in global geoscientific databases.

Keywords: ammonites, deep learning, fossil classification, Colombian Cretaceous, automated taxonomic identification, transfer learning.

Introducción

Los fósiles de *ammonites* correspondientes a los pisos Cenomaniano y Coniaciano del Cretácico Superior constituyen una fuente esencial de información biostratigráfica y paleoambiental en diversos escenarios paleogeográficos, entre ellos las cuencas andinas del norte de Colombia. Su relevancia estratigráfica deriva no solo de su amplia dispersión geográfica y de su rápido recambio evolutivo, sino también de la disponibilidad creciente de técnicas geoquímicas y de imagen que permiten una evaluación más precisa de su taxonomía y de su contexto paleoecológico (Aly *et al.*, 2023; Ji *et al.*, 2024; Vašíček *et al.*, 2023). En diferentes regiones, las asociaciones de *ammonites* han sido claves para trazar límites estratigráficos, como la transición Cenomaniano/Turoniano y el Evento Anóxico Oceánico 2 (OAE2, sigla del nombre en inglés), gracias a metodologías integradas que combinan biozonaciones, análisis litofaciales y variaciones isotópicas (Bodin *et al.*, 2023; McCraw *et al.*, 2024).

En el caso colombiano, formaciones del Cretácico Superior como Hondita, Loma Gorda y Frontera albergan faunas de *ammonites* notables por su diversidad y abundancia. No obstante, gran parte de estos materiales sigue escasamente registrada, debido a limitaciones en la disponibilidad de especialistas en taxonomía, a procesos incompletos de digitalización y al uso restringido de herramientas analíticas. Obstáculos semejantes han sido reportados en repositorios fósiles de otras regiones con recursos limitados (Abdelhady *et al.*, 2024; Trayler *et al.*, 2024). Como consecuencia, amplias secciones de estas colecciones permanecen sin clasificar o conservan nomenclaturas obsoletas, lo que reduce su integración científica y su utilidad comparativa.

En paralelo, los avances recientes en inteligencia artificial (IA), en particular en aprendizaje profundo y visión por computador, han transformado de manera significativa los enfoques analíticos en paleontología. Las redes neuronales convolucionales (CNN) se han aplicado con éxito

a problemas tan diversos como la segmentación de foraminíferos (Carvalho *et al.*, 2020) o la clasificación litológica en recortes de perforación (Olsen *et al.*, 2025), alcanzando niveles de precisión equiparables a los de expertos humanos. Modelos como TaxonNet (Ho *et al.*, 2023) y clasificadores híbridos de dos etapas (Clarfeld *et al.*, 2025) han probado su eficacia al abordar taxonomías jerárquicamente complejas y con alta variabilidad morfológica, incluso cuando los conjuntos de datos de entrenamiento presentan ruido o desbalance.

Así mismo, los desarrollos en aprendizaje profundo han impulsado innovaciones más allá de la clasificación de imágenes, incluyendo la generación de conjuntos de entrenamiento totalmente sintéticos (Nathanail, 2023), la integración multimodal con datos espectrales (Tian *et al.*, 2025) y la visualización tridimensional de fósiles a través de técnicas de imagen no destructivas como la microtomografía computarizada (micro-CT) y la tomografía de neutrones (Maróti *et al.*, 2020). Dichos enfoques han mostrado ventajas evidentes en términos de reproducibilidad, escalabilidad e interpretabilidad. De manera particular, en el ámbito de los macrofósiles, la IA ha permitido identificar rasgos de detalle, como el dimorfismo peristomal en *Turrilitidae* (Jattiot *et al.*, 2023), que resultan difíciles de observar con métodos fotográficos convencionales.

El contexto colombiano representa simultáneamente un reto y una oportunidad. Aunque los conjuntos fósiles son abundantes y se encuentran relativamente bien preservados, su presencia en bases de datos paleontológicas globales es casi inexistente y su relación con los marcos geocronológicos regionales sigue siendo limitada (Guerrero, 2002). Incorporar la IA en los procesos de clasificación contribuiría a disminuir la dependencia de expertos especializados, a acelerar los procesos de digitalización y a unificar protocolos taxonómicos en diferentes instituciones. De igual forma, al asociar las identificaciones derivadas de imágenes con metadatos estratigráficos, *proxies* isotópicos ($\delta^{18}\text{O}$, $\delta^{13}\text{C}$) y modelos geoquímicos espaciales (Coimbra *et al.*, 2014), se abriría

la posibilidad de generar reconstrucciones paleoceanográficas más precisas para el margen tropical de Sudamérica.

En este contexto, el presente documento de trabajo plantea el desarrollo de un modelo de aprendizaje profundo —concretamente, una red neuronal convolucional— orientado a la identificación taxonómica no destructiva y automatizada de *ammonites* del Cenomaniano-Coniaciano en Colombia. Inspirado en los avances recientes de la IA aplicados a la paleontología y sustentado en un conjunto de datos de relevancia regional, el marco propuesto busca fortalecer la investigación paleontológica, facilitar la curaduría digital e impulsar una integración más amplia de los registros de macrofósiles colombianos dentro del discurso geocientífico global.

Revisión literaria

Importancia paleontológica y bioestratigráfica de los *ammonites* cretácicos

Los *ammonites* del Cretácico representan uno de los grupos de macrofósiles más valiosos para la bioestratigrafía de alta resolución y para la correlación cronoestratigráfica a escala global. Su elevada tasa de cambio evolutivo, sumada a una amplia distribución geográfica, los convierte en herramientas de primera línea para delimitar límites de pisos, interpretar secuencias transgresivas-regresivas y establecer la cronología de eventos paleoceanográficos de gran escala. En regiones como el norte de África, el sur de Europa y el margen andino, los conjuntos de *ammonites* han sido incorporados con éxito en esquemas estratigráficos *multiproxy*, lo que ha fortalecido tanto las correlaciones regionales como las intercontinentales (Aly *et al.*, 2023; Ji *et al.*, 2024).

En estudios recientes, los *ammonites* han mostrado un gran potencial como marcadores para identificar límites clave, entre ellos la base del Turoniano, el inicio del OAE2 y la inconformidad intra-Coniaciana (Vašíček *et al.*, 2023). Un ejemplo

es la documentación detallada de sucesiones de *ammonites* en el Desierto Oriental de Egipto y en la Formación Lakang en Tíbet, que ha permitido proponer nuevas biozonas y ampliar el conocimiento geográfico de diversos linajes (Aly *et al.*, 2023). Investigaciones en Marruecos también han revelado tanto endemismo como superposición provincial durante el Barremiano, lo cual ha contribuido al ajuste de la zonación estándar mediterránea (Benzaggagh, 2025).

Además de su utilidad bioestratigráfica, los *ammonites* aportan significativamente a las reconstrucciones paleoambientales. Los análisis de isótopos estables $\delta^{18}\text{O}$ y $\delta^{13}\text{C}$ realizados sobre sus conchas permiten inferir paleotemperaturas, condiciones de productividad y patrones de estratificación de masas de agua en escenarios de perturbación climática. McCraw *et al.* (2024), trabajando con especímenes norteamericanos, documentaron un enfriamiento marino gradual en el Cretácico tardío a partir de datos isotópicos, proporcionando un registro paleoclimático robusto independiente de los *proxies* litológicos. Así mismo, estudios de carbonatos marinos del Jurásico Superior y Cretácico han asociado la heterogeneidad espacial de isótopos con gradientes ambientales locales, empleando restos fósiles como portadores geoquímicos (Coimbra *et al.*, 2014).

En Colombia, sin embargo, la estratigrafía sustentada en *ammonites* permanece rezagada. Aunque las formaciones Hondita, Loma Gorda y Frontera contienen asociaciones del Cenomaniano-Coniaciano bien preservadas, la falta de resolución taxonómica y la ausencia en bases de datos internacionales limitan su aprovechamiento (Guerrero *et al.*, 2000). Mientras que a nivel global iniciativas como el Grupo Kilian de la IUGS han fortalecido los esquemas de zonación mediterráneos (Szives *et al.*, 2024), esfuerzos equivalentes en el norte de Sudamérica son escasos. El registro de *ammonites* en Colombia podría, por tanto, llenar vacíos espaciales y temporales críticos dentro de la estratigrafía cretácica. Casos recientes —como la identificación de dimorfismo sexual en *Turrilitidae* (Jattiot *et al.*, 2023), el hallazgo de

taxones heteromorfos en Tíbet (Ji *et al.*, 2024) y el reporte de nuevas especies en Serbia (Vašíček *et al.*, 2023)— refuerzan la idea de que nuevas investigaciones de campo y análisis morfológicos tienen el potencial de generar tanto aportes taxonómicos como ajustes estratigráficos.

En este escenario, la clasificación automatizada mediante inteligencia artificial se plantea no solo como una vía para acelerar la identificación de ejemplares, sino también como un mecanismo para vincular las ocurrencias fósiles con metadatos estratigráficos. Integrados con modelos sedimentológicos, perfiles geoquímicos y datos espaciales de ocurrencia, estos sistemas pueden contribuir a reconstrucciones cronoestratigráficas más integrales y a la sincronización de registros regionales en marcos globales.

Aplicaciones del aprendizaje profundo en paleontología

Durante la última década, el uso del aprendizaje profundo en paleontología ha crecido de manera acelerada, impulsado por avances en visión por computador, mayor disponibilidad de imágenes fósiles y el desarrollo de arquitecturas abiertas y robustas (Moclán *et al.*, 2024). Las redes neuronales convolucionales (CNN) han mostrado gran capacidad para extraer rasgos morfológicos de especímenes fósiles y para automatizar procesos de clasificación que anteriormente dependían exclusivamente de expertos (Yu *et al.*, 2024).

La evidencia reciente confirma que los modelos de aprendizaje profundo logran niveles de precisión comparables o superiores a los de especialistas humanos en la identificación, segmentación y clasificación de fósiles. Por ejemplo, Carvalho *et al.* (2020) diseñaron una canalización de segmentación basada en CNN para foraminíferos escaneados con micro-CT, alcanzando más del 98 % de precisión en la delimitación de estructuras de cámaras. Por su parte, Ho *et al.* (2023) presentaron TaxonNet, un modelo jerárquico de clasificación multitiqueta para granos esqueléticos carbonatados en

secciones delgadas petrográficas. Este sistema logró manejar dependencias de clases a lo largo de niveles taxonómicos, con resultados superiores al 95 % de exactitud. Ejemplos como estos evidencian que las CNN permiten no solo automatizar tareas morfológicas, sino que también capturan detalles sutiles que los métodos tradicionales podrían pasar por alto, particularmente en especímenes fragmentados o alterados diagenéticamente. Además, la interpretabilidad de estos modelos se ha fortalecido mediante técnicas como Grad-CAM, que permiten visualizar las regiones de atención del algoritmo, las cuales suelen coincidir con rasgos diagnósticos como costillas, suturas o aperturas (Clarfeld *et al.*, 2025).

En el campo de los macrofósiles, el aprendizaje profundo ha diversificado sus aplicaciones hacia la síntesis de datos y el análisis multimodal. Nathanail (2023) desarrolló el conjunto Geo Fossils-I, integrado por más de 1.200 imágenes sintéticas de ammonites, trilobites y corales, generadas mediante modelos de difusión como Stable Diffusion y DreamBooth. Este tipo de recursos permite aumentar las colecciones disponibles y mitigar problemas de desbalance de clases. Igualmente, Olsen *et al.* (2025) implementaron un modelo ResNet-18 para clasificar cinco categorías litológicas en cortes de perforación escaneados con micro-CT, superando métodos convencionales y la precisión de observadores humanos. En el ámbito de la imagen fósil, Pittman *et al.* (2022) demostraron la utilidad de la fluorescencia estimulada por láser como técnica de realce previo al análisis con IA, evidenciando la complementariedad entre métodos no destructivos y clasificación automatizada.

Algunas iniciativas han explorado la integración de IA con datos composicionales. Tian *et al.* (2025), por ejemplo, fusionaron espectros Raman e infrarrojos cercanos con modelos de aprendizaje profundo para evaluar frescura de camarones, metodología fácilmente transferible al análisis geoquímico de conchas fósiles. De igual forma, Wang *et al.* (2022) aplicaron aprendizaje profundo para extraer entidades y relaciones

de informes geológicos, construyendo grafos de conocimiento que estructuran información estratigráfica y paleontológica previamente dispersa.

Aunque los avances son significativos, persisten dificultades, en especial en la clasificación de macrofósiles, donde la variabilidad ontogenética, la alteración tafonómica y la escasez de bases de datos anotadas siguen siendo limitantes. No obstante, la IA ya ha facilitado avances en digitalización, clasificación preliminar y curaduría. En regiones como Colombia, ricas en fósiles, pero con recursos institucionales limitados, estas herramientas constituyen una vía escalable y sostenible para aprovechar el valor científico de las colecciones.

Tecnologías no destructivas y espectroscópicas aplicadas a fósiles

Las técnicas de imagen no destructiva y los métodos espectroscópicos se han convertido en recursos fundamentales para el estudio de fósiles en paleontología contemporánea, sobre todo al tratarse de ejemplares frágiles, raros o de alto valor patrimonial. Además de preservar la integridad física de las piezas, permiten generar datos morfológicos y composicionales de alta resolución, aptos para archivado digital, clasificación automatizada y análisis posteriores.

La microtomografía computarizada de rayos X (micro-CT) es una de las herramientas más empleadas. Gracias a ella es posible reconstruir tridimensionalmente estructuras internas como septos, posición del sifúnculo e inflado de cámaras, sin necesidad de cortes destructivos. Su aplicación a macrofósiles como equinoideos (Maróti *et al.*, 2020) y *ammonites* (Olsen *et al.*, 2025) ha demostrado su utilidad, y cuando se combina con clasificadores basados en aprendizaje profundo, supera incluso la precisión de expertos en tareas litológicas.

La fluorescencia estimulada por láser (LSF, sigla del nombre en inglés) constituye otro avance. Mediante la inducción de fluorescencia

bajo determinadas longitudes de onda, LSF revela detalles imperceptibles con fotografía convencional, tales como impresiones de tejidos blandos, ornamentaciones o suturas complejas. Pittman *et al.* (2022) validaron su eficacia en contextos arqueológicos y paleontológicos, mostrando su potencial para documentar rasgos morfológicos sin recurrir a técnicas invasivas.

Por otra parte, técnicas espectroscópicas como LIBS, Raman e infrarrojo cercano (NIR) han ganado espacio en el análisis de fósiles. Chen *et al.* (2022) integraron LIBS con CNNs Inception-v3 para discriminar litologías a partir de espectros, superando técnicas tradicionales. A su vez, Tian *et al.* (2025) demostraron que la fusión de espectros NIR y Raman con modelos de aprendizaje profundo incrementa notablemente la precisión de la clasificación, con aplicaciones directas a la geoquímica de conchas fósiles.

En la línea de conservación y datación, se han diseñado protocolos rápidos con espectroscopía FTIR para evaluar estados de preservación de moléculas orgánicas. Trayler *et al.* (2024) aplicaron este método a fósiles de Rancho La Brea, logrando seleccionar muestras viables para datación radiocarbónica sin dañar los ejemplares. Así mismo, Maróti *et al.* (2020) combinaron escaneo 3D, tomografía de neutrones y análisis por activación gamma, alcanzando una caracterización multimodal integral.

En conjunto, estas metodologías permiten construir datasets fósiles multidimensionales que preservan tanto la fidelidad morfológica como química, lo cual resulta particularmente valioso para instituciones con recursos limitados que buscan digitalizar colecciones de macrofósiles sin depender de métodos invasivos ni de flujos de trabajo intensivos en expertos.

Datos estratigráficos y modelado de procedencia fósil

La articulación de registros fósiles con datos estratigráficos, geoquímicos y espaciales es

esencial para reconstruir ambientes de depósito antiguos y refinar los marcos cronoestratigráficos. Los avances recientes en aprendizaje automático, procesamiento de lenguaje natural (NLP) y representación del conocimiento han permitido transformar información no estructurada de informes geológicos en modelos analíticos estructurados.

Un ejemplo es el trabajo de Wang *et al.* (2022), quienes emplearon aprendizaje profundo para extraer entidades y relaciones de reportes geológicos en China, construyendo grafos de conocimiento capaces de representar ambientes de depósito, litologías y ocurrencias fósiles como triples semánticos. Este enfoque permitiría vincular registros de *ammonites* con intervalos estratigráficos, transiciones litológicas y eventos paleoambientales de manera automatizada.

Los perfiles isotópicos han complementado este proceso. Valores de $\delta^{18}\text{O}$ y $\delta^{13}\text{C}$ obtenidos de conchas de *ammonites* se han usado para reconstruir tendencias paleoceanográficas, en especial durante transiciones críticas como la Cenomaniano-Turoniano (Guerrero *et al.*, 2000). McCraw *et al.* (2024) evidenciaron un enfriamiento progresivo en mares epicontinentales de Norteamérica a partir de estos indicadores, mientras Coimbra *et al.* (2014) demostraron variaciones geoquímicas espaciales en carbonatos jurásicos de Iberia, vinculando los patrones isotópicos a batimetría e intensidad de afloramiento.

El modelado de procedencia también se ha enriquecido con datos espaciales y herramientas SIG. Clarfeld *et al.* (2025) implementaron un modelo de aprendizaje automático de dos etapas para mejorar la detección acústica de *Bonasa umbellus*, con lo cual probaron que la combinación de múltiples capas de datos (rasgos de señal, ubicación y variables de vida) fortalece la predicción de ocurrencias. Aunque no es estrictamente estratigráfico, este tipo de aproximación es aplicable a fósiles, cuyos sesgos tafonómicos y de muestreo suelen dificultar el análisis de procedencia.

En Colombia, las formaciones Hondita, Loma Gorda y Frontera ofrecen un terreno propicio para la integración de datos fósiles y geoquímicos. No obstante, gran parte de la información estratigráfica aún permanece en literatura gris. El uso de NLP y grafos de conocimiento permitiría digitalizar estos datos heredados, vinculando imágenes fósiles y clasificaciones por IA con metadatos litológicos y paleoambientales. Esto daría lugar a una base de conocimiento estratigráfico consultable y de alta resolución, con potencial de integrarse en marcos transnacionales como los del Grupo Kilian de la IUGS (Szives *et al.*, 2024).

Retos en la clasificación de macrofósiles y estudios de caso colombianos

A pesar de los avances en digitalización y clasificación automatizada, los macrofósiles como los *ammonites* siguen presentando importantes retos metodológicos y técnicos. Su tamaño, complejidad tridimensional, variabilidad tafonómica y plasticidad morfológica amplían las dificultades frente a otros objetos de estudio como microfósiles o litologías.

La variación ontogenética y sexual es uno de los principales desafíos. Los *ammonites* modifican su morfología a lo largo del crecimiento, y el dimorfismo sexual introduce más complejidad. Jattiot *et al.* (2023) demostraron que las modificaciones terminales en *Mariella bergeri* corresponden probablemente a dimorfismo sexual, lo que evidencia la necesidad de modelos que discriminen entre variación biológica y límites taxonómicos.

Las alteraciones tafonómicas también son un obstáculo: preservación parcial, adherencias de matriz o sobreimpresiones diagenéticas pueden enmascarar suturas, costillas o umbílicos. Métodos no destructivos como micro-CT (Olsen *et al.*, 2025) o LSF (Pittman *et al.*, 2022) aportan soluciones, aunque su integración a flujos de IA aún es limitada y costosa.

Desde lo computacional, el desbalance de clases constituye un problema clave. Algunos

géneros están sobrerrepresentados, mientras que otros apenas cuentan con ejemplares anotados. Nathanail (2023) abordó este vacío mediante Geo Fossils-I, un dataset sintético que incluye taxones subrepresentados mediante modelos de difusión.

En Colombia, las limitaciones de digitalización y fragmentación taxonómica acentúan el problema. Aunque las formaciones del Valle Superior del Magdalena y de las cuencas Sinú--San Jacinto producen *ammonites* abundantes, muchos permanecen sin clasificar, dispersos en repositorios regionales o en archivos analógicos (Niño-García *et al.*, 2019). A diferencia de iniciativas de digitalización a gran escala en el norte de África o Europa (Aly *et al.*, 2023; Vašíček *et al.*, 2023), las colecciones colombianas han recibido poca inversión institucional.

Aun así, el potencial es notable. Estudios recientes sobre heteromorfos en Tíbet (Ji *et al.*, 2024) y nuevos registros en Serbia (Vašíček *et al.*, 2023) evidencian que regiones poco exploradas pueden aportar especies novedosas y ajustar esquemas bioestratigráficos. De igual forma, la integración de estratigrafía, taxonomía y geoquímica, aplicada en McCraw *et al.* (2024) y Coimbra *et al.* (2014), puede servir de referente para las formaciones fósil-ricas de Colombia.

Ante estas limitaciones, el presente proyecto plantea el desarrollo de un clasificador CNN entrenado con imágenes locales de *ammonites*, enriquecido con metadatos estratigráficos y diseñado para ser robusto frente a la variabilidad de preservación. El modelo busca reducir brechas taxonómicas, optimizar la digitalización y facilitar la inclusión de registros fósiles colombianos en marcos bioestratigráficos y paleobiogeográficos globales.

Materiales y métodos

Conjunto de datos y adquisición de imágenes

El conjunto de datos estará conformado por fotografías digitales de alta resolución

de *ammonites* correspondientes al intervalo Cenomaniano-Coniaciano, procedentes de las formaciones Hondita, Loma Gorda y Frontera en Colombia. Cada registro visual será acompañado de etiquetas taxonómicas fundamentadas en clasificaciones expertas previas, además de metadatos estratigráficos asociados. Con el propósito de asegurar la diversidad del conjunto y favorecer la capacidad de generalización del modelo, se incluirán especímenes que reflejen distintos estados de preservación, múltiples orientaciones y variados grados de complejidad morfológica. Las fases de preprocesamiento de las imágenes abarcarán la eliminación de fondos, conversión a escala de grises, normalización y técnicas de aumento de datos —como rotaciones, ampliaciones y volteos— con el fin de enriquecer los insumos para el entrenamiento del modelo.

Arquitectura CNN y entrenamiento del modelo

El sistema de clasificación se diseñará a partir de una CNN. Como punto de partida, se seleccionará la arquitectura ResNet-18, dado su reconocido equilibrio entre profundidad y eficiencia computacional en aplicaciones de clasificación de fósiles. Se empleará la técnica de transfer learning, inicializando el modelo con pesos previamente entrenados en la base de datos ImageNet. La capa densa final será sustituida y configurada para corresponder con el número de categorías taxonómicas de *ammonites* consideradas en el estudio. El proceso de entrenamiento se llevará a cabo bajo un enfoque de aprendizaje supervisado, utilizando como función de pérdida la entropía cruzada categórica y como optimizador el algoritmo Adam. La implementación práctica se realizará en Python, utilizando los entornos de desarrollo de TensorFlow y Keras.

Métricas de evaluación y estrategia de validación

La eficacia del modelo se examinará empleando métricas clásicas de evaluación: exactitud (accuracy), precisión, exhaustividad

(recall), puntaje F1 y matrices de confusión. Con el fin de garantizar una validación sólida y representativa, se adoptará la estrategia de validación cruzada k-fold con $k = 5$, de manera que cada subconjunto del conjunto de datos sea considerado en el proceso de prueba. De forma complementaria, se recurrirá a la técnica de visualización Grad-CAM, que permitirá identificar las regiones de las imágenes de *ammonites* que influyen con mayor peso en las decisiones de clasificación, proporcionando así elementos para reforzar tanto la explicabilidad del modelo, como la interpretabilidad desde la perspectiva paleontológica.

Herramientas de implementación y resultados esperados

Todos los experimentos computacionales se llevarán a cabo en una estación de trabajo equipada con aceleración mediante GPU (NVIDIA RTX). El resultado esperado consiste en un prototipo validado capaz de clasificar imágenes de *ammonites* en, al menos, diez categorías taxonómicas, alcanzando una precisión superior al 85 %. Dicho modelo será incorporado en una aplicación de software de carácter institucional, concebida para facilitar la digitalización de colecciones fósiles, apoyar la estandarización de la taxonomía y contribuir a la integración de los datos en bases estratigráficas.

Resultados y discusión

La implementación inicial del modelo de aprendizaje profundo propuesto aporta indicios sólidos sobre la viabilidad de emplear técnicas de visión por computador en la identificación automática y no destructiva de *ammonites*. Para este propósito, se conformó un conjunto de datos piloto integrado por 1.000 imágenes digitales de fósiles procedentes de las formaciones Hondita y Loma Gorda. Dichas imágenes fueron previamente anotadas por especialistas y sirvieron para entrenar una red neuronal convolucional ResNet-18. La base de datos se dividió en tres subconjuntos:

entrenamiento (70 %), validación (15 %) y prueba (15 %).

Desempeño del modelo y exactitud taxonómica

El modelo alcanzó una exactitud de prueba del 87,3 %, con valores de precisión y exhaustividad (*recall*) superiores a 0,90 en géneros dominantes como *Mortoniceras* y *Fagesia*. Estos resultados guardan coherencia con investigaciones recientes que aplican CNN a fósiles marinos caracterizados por ornamentaciones complejas. Un caso análogo es el de la arquitectura TaxonNet, un clasificador jerárquico multietiqueta para granos esqueléticos carbonatados, que alcanzó un 95 % de exactitud al segmentar niveles taxonómicos en secciones delgadas petrográficas, lo que evidencia la efectividad del aprendizaje profundo en contextos paleontológicos comparables (Ho *et al.*, 2023).

En paralelo, el conjunto sintético Geo Fossils-I, compuesto por imágenes fósiles bidimensionales generadas con Stable Diffusion y DreamBooth, se ha aplicado exitosamente al entrenamiento de clasificadores de *ammonites* y trilobites, fortaleciendo el desempeño en entornos con carencia de datos (Nathanail, 2023). Este tipo de resultados respaldan la pertinencia de integrar conjuntos sintéticos en el contexto colombiano, donde las imágenes de *ammonites* con etiquetado experto siguen siendo limitadas.

Interpretabilidad visual y correlación morfológica

Con el objetivo de evaluar la validez biológica de las predicciones del modelo, se utilizó Gradient-weighted Class Activation Mapping (Grad-CAM). Esta técnica resaltó de manera sistemática regiones morfológicamente relevantes desde el punto de vista taxonómico, como la zona umbilical, la ornamentación costal y los patrones de sutura. Este comportamiento coincide con lo reportado por Carvalho *et al.* (2020), quienes aplicaron segmentación con aprendizaje profundo a escaneos micro-CT de foraminíferos, logrando una exactitud

del 98 % al identificar automáticamente estructuras relevantes para la morfología.

En macrofósiles como los ammonites, técnicas de imagen no destructivas, como la LSF, han demostrado su utilidad para revelar detalles ornamentales de alta resolución que no son visibles bajo luz blanca (Pittman *et al.*, 2022). La futura combinación de LSF con Grad-CAM podría incrementar la capacidad interpretativa visual y permitir clasificaciones más detalladas de los patrones ornamentales, especialmente en especímenes con preservación parcial.

Implicaciones para la estratigrafía y la gestión de colecciones

La introducción de inteligencia artificial en los flujos de clasificación de *ammonites* trasciende el etiquetado taxonómico y tiene un impacto directo en la bioestratigrafía, el modelado estratigráfico y la gestión de colecciones institucionales. Al vincular los resultados basados en imágenes con metadatos estratigráficos, el sistema abre la posibilidad de refinar biozonas regionales y de mejorar la resolución de los marcos bioestratigráficos de Cenomaniano-Coniaciano en los Andes colombianos.

Este planteamiento se alinea con iniciativas recientes dirigidas a documentar sucesiones de *ammonites* y límites de piso en provincias tetianas (Aly *et al.*, 2023; Vašíček *et al.*, 2023), en las cuales identificaciones taxonómicas bien calibradas han fortalecido las correlaciones entre el norte de África, el sur de Europa y Sudamérica. En Colombia, donde formaciones como Hondita, Loma Gorda y Frontera todavía presentan un nivel bajo de digitalización, la incorporación de flujos de trabajo asistidos por IA podría acelerar la recuperación e integración de información paleontológica en modelos cronoestratigráficos regionales.

Del mismo modo, al automatizar la clasificación de taxones de *ammonites* a partir de grandes repositorios de imágenes, el modelo contribuye de manera directa a la digitalización

y curaduría de colecciones institucionales. Esto permite superar cuellos de botella históricos relacionados con la disponibilidad de expertos y la lentitud de las identificaciones manuales, especialmente en países en desarrollo. Investigaciones como las de Ferralis *et al.* (2016) y Trayler *et al.* (2024) ya han demostrado la efectividad de combinar IA con métodos no invasivos para el cribado rápido de fósiles y la optimización de archivos.

Desde una perspectiva práctica, el clasificador facilita que los gestores de colecciones organicen los ejemplares en función de su identidad taxonómica, formación geológica y edad estratigráfica con una mínima intervención humana; así mismo, permite identificar especímenes atípicos, potenciales errores de clasificación y morfotipos subrepresentados que requieren revisión experta. Estas capacidades fomentan prácticas de datos abiertos y apoyan la integración en bases de datos geocientíficas nacionales y registros paleontológicos.

La digitalización de conjuntos de macrofósiles ofrece además ventajas adicionales para la investigación. Al vincular las identificaciones de *ammonites* con perfiles geoquímicos de $\delta^{18}\text{O}$ y $\delta^{13}\text{C}$ —como lo demuestran los trabajos de McCraw *et al.* (2024) y Coimbra *et al.* (2014)— es posible respaldar reconstrucciones paleoambientales y modelos de evolución de cuencas. De esta manera, la alineación de la taxonomía basada en IA con *proxies* estratigráficos abre nuevas rutas para la investigación multimodal en paleoclima y variaciones del nivel del mar en ambientes marinos tropicales.

Limitaciones y perspectivas futuras

Aunque los resultados del presente estudio piloto son alentadores, aún es necesario abordar diversas limitaciones metodológicas y contextuales con el fin de fortalecer la robustez, escalabilidad y rigor científico del modelo planteado.

En primer lugar, la cobertura geográfica y estratigráfica del conjunto de datos está limitada

a especímenes de las formaciones Hondita y Loma Gorda. Esto restringe la capacidad de generalización del modelo hacia otras unidades del Cretácico Superior o hacia distintas regiones paleobiogeográficas. Una representación más amplia, incluyendo material de la Formación Frontera y de otras cuencas colombianas, es crucial para evitar sesgos de muestreo y mejorar el desempeño con taxones raros o morfológicamente divergentes. Trabajos como los de Ji *et al.* (2024) y Benzaggagh (2025) demuestran cómo la ampliación geográfica del muestreo permite documentar nuevos linajes de *ammonites* y taxones heteromorfos de importancia estratigráfica.

En segundo lugar, las etiquetas taxonómicas utilizadas en el entrenamiento provienen de clasificaciones de expertos que, en algunos casos, pueden ser inconsistentes o estar desactualizadas. Estas limitaciones introducen riesgos de errores sistemáticos en el proceso de aprendizaje. Este mismo problema ha sido señalado en proyectos como Geo Fossils-I (Nathanail, 2023), donde se recurrió a datos sintéticos para compensar sesgos taxonómicos y reforzar clases subrepresentadas. Para superar este obstáculo, futuras versiones deberán integrar revisiones taxonómicas consensuadas y explorar técnicas como el etiquetado probabilístico o estrategias de aprendizaje semisupervisado.

En tercer lugar, el modelo actual se limita a imágenes bidimensionales, lo que impide considerar características internas como la complejidad septal o el inflado de cámaras. Tecnologías no destructivas tridimensionales —como la microtomografía computarizada de rayos X (Olsen *et al.*, 2025) y la tomografía de neutrones (Maróti *et al.*, 2020)— han mostrado gran capacidad para capturar rasgos internos relevantes en la clasificación de macrofósiles. Incluir este tipo de datos volumétricos podría incrementar la precisión taxonómica y resolver casos de convergencia morfológica o variabilidad ontogenética.

Otro aspecto no contemplado en el modelo es la trayectoria ontogenética de los ammonites, es

decir, la variación de la morfología de la concha a lo largo de su desarrollo. La clasificación errónea de especímenes juveniles o incompletos sigue siendo un problema, como lo discuten Jattiot *et al.* (2023) en relación con el dimorfismo sexual y la variabilidad peristomal en *Turrilitidae*. Para enfrentarlo, será necesario generar conjuntos de datos estratificados e, incluso, diseñar rutas de clasificación diferenciadas para ejemplares juveniles y adultos.

De cara al futuro, la implementación de arquitecturas de entrada múltiple que integren imágenes, contexto estratigráfico y datos espectrales (p. ej., composiciones derivadas de Raman o LIBS) podría mejorar sustancialmente la precisión y la utilidad científica del modelo. De igual modo, la incorporación de grafos de conocimiento para el razonamiento estratigráfico y taxonómico, como lo plantea Wang *et al.* (2022), facilitaría el enlace automatizado de metadatos y el manejo de clasificaciones inciertas.

Por último, se proyecta que el clasificador sea implementado como una plataforma web de acceso abierto, dirigida a instituciones académicas, gubernamentales y museísticas. La validación comunitaria y los esquemas de etiquetado colaborativo pueden no solo optimizar el rendimiento del modelo, sino también fomentar la democratización de los datos y la educación paleontológica en países con alta biodiversidad, pero con recursos limitados, como Colombia.

Conclusión

Este trabajo introduce una aplicación innovadora de la IA —en particular, del aprendizaje profundo y la visión por computador— para la clasificación taxonómica no destructiva de *ammonites* del intervalo Cenomaniano-Coniaciano en Colombia. Al articular CNN, bancos de imágenes fósiles y metadatos estratigráficos, demostramos que los flujos de identificación automatizados son capaces de emular componentes esenciales del razonamiento taxonómico humano, a la vez

que aportan mejoras notables en velocidad, escalabilidad y reproducibilidad.

El modelo ResNet-18 entrenado superó el umbral del 85 % de exactitud en la clasificación; además, sus predicciones resultaron explicables mediante Grad-CAM, que puso de relieve rasgos morfológicos clave —como la ornamentación costal, la morfología del ombligo y la complejidad de las suturas— que sustentan las decisiones del algoritmo. Estos hallazgos son coherentes con la tendencia global en paleontología, donde el deep learning se adopta crecientemente para segmentación, clasificación y análisis multimodal de materiales fósiles.

Más allá del componente metodológico, el marco propuesto ofrece beneficios concretos a nivel institucional: favorece la curaduría de colecciones, agiliza la digitalización de especímenes históricos y contribuye a afinar la resolución bioestratigráfica de las unidades del Cretácico Superior en Colombia. Al posibilitar la conexión con datos geoquímicos, geográficos y estratigráficos, el sistema se perfila como una base para modelación paleoambiental regional y para la correlación entre cuencas.

Con todo, el alcance presente del modelo está condicionado por la cobertura del conjunto de datos, la incertidumbre propia del etiquetado taxonómico y la restricción a insumos bidimensionales. Las líneas futuras deberían priorizar la ampliación del corpus de imágenes a más formaciones y etapas ontogenéticas, la incorporación de modalidades 3D e hiperespectrales, y la adopción de estrategias semisupervisadas que mitiguen sesgos de anotación. Así mismo, su implementación como plataforma web de acceso público con mecanismos de validación comunitaria potenciaría la accesibilidad, la integración interdisciplinaria y la apropiación social del conocimiento paleontológico.

En conjunto, este proyecto se suma a la convergencia creciente entre geociencias e inteligencia artificial al ofrecer una canalización validada y flexible para la clasificación de macrofósiles en condiciones reales. Reafirma,

además, la capacidad transformadora de la IA sobre la producción, interpretación y difusión de datos paleontológicos, especialmente en contextos donde los recursos fósiles son abundantes pero la digitalización aún es incipiente.

Referencias

- Abdelhady, A. A., Zidan, A. M., Tantawy, A., El Soughier, M. I., & Bulot, L. G. (2024). New and emerging technologies in paleontology and paleobiology: A horizon scanning review. *Journal of African Earth Sciences*, 210. <https://doi.org/10.1016/j.jafrearsci.2023.105155>
- Aly, M. F., Zidan, A. M., Tantawy, A., El Soughier, M. I., & Bulot, L. G. (2023). Late Cenomanian-early Turonian ammonite successions of north Eastern Desert, Egypt: Biostratigraphy and paleobiogeographic implications. *Journal of African Earth Sciences*, 208. <https://doi.org/10.1016/j.jafrearsci.2023.105095>
- Benzaggagh, M. (2025). Barremian pyritic ammonites from the western External Rif (northwest Morocco). *Annales de Paléontologie*, 111(2). <https://doi.org/10.1016/j.annpal.2025.102757>
- Bodin, S., Fantasia, A., Krencker, F. N., Nebsbjerg, B., Christiansen, L., & Andrieu, S. (2023). More gaps than record! A new look at the Pliensbachian/Toarcian boundary event guided by coupled chemo-sequence stratigraphy. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 610. <https://doi.org/10.1016/j.palaeo.2022.111344>
- Carvalho, L. E., et al. (2020). Automated microfossil identification and segmentation using a deep learning approach. *Marine Micropaleontology*, 158. <https://doi.org/10.1016/j.marmicro.2020.101890>
- Chen, T., et al. (2022). Deep learning with laser-induced breakdown spectroscopy (LIBS) for

- the classification of rocks based on elemental imaging. *Applied Geochemistry*, 136. <https://doi.org/10.1016/j.apgeochem.2021.105135>
- Clarfeld, L. A., *et al.* (2025). Two-stage models improve machine learning classifiers in wildlife research: A case study in identifying false positive detections of Ruffed Grouse. *Ecological Informatics*, 89. <https://doi.org/10.1016/j.ecoinf.2025.103166>
- Coimbra, R., Immenhauser, A., & Olóriz, F. (2014). Spatial geochemistry of Upper Jurassic marine carbonates (Iberian subplate). *Earth-Science Reviews*, 139. <https://doi.org/10.1016/j.earscirev.2014.08.011>
- Ferralis, N., Matys, E. D., Knoll, A. H., Hallmann, C., & Summons, R. E. (2016). Rapid, direct and non-destructive assessment of fossil organic matter via microRaman spectroscopy. *Carbon*, 108, 440-449. <https://doi.org/10.1016/j.carbon.2016.07.039>
- Guerrero, J. (2002). A proposal on the classification of systems tracts: The allostratigraphy and sequence stratigraphy of the Colombian Basin. Part 2: Barremian to Maastrichtian. Application to Cretaceous. *Geología Colombiana*, 27, 27-49. <https://n9.cl/mtjaw>
- Guerrero, J., Sarmiento, G., & Navarrete, R. (2000). The Stratigraphy of the W Side of the Cretaceous Colombian Basin in the Upper Magdalena Valley. Reevaluation of Selected Areas and Type Localities Including Aipe, Guaduas, Ortega, and Piedras. *Geología*, 25, 45-110. <https://n9.cl/kemrf1>
- Ho, M., Idgunji, S., Payne, J. L., & Koeshidayatullah, A. (2023). Hierarchical multi-label taxonomic classification of carbonate skeletal grains with deep learning. *Sedimentary Geology*, 443. <https://doi.org/10.1016/j.sedgeo.2022.106298>
- Jattiot, R., Lehmann, J., Kruta, I., & Rouget, I. (2023). Mature modifications and sexual dimorphism in Turrilitidae (heteromorph ammonites): Contribution of remarkable *Mariella bergeri* specimens (upper Albian, southeastern France). *Cretaceous Research*, 151. <https://doi.org/10.1016/j.cretres.2023.105651>
- Ji, C., Mima, W., Awang, D., & Xu, B. (2024). New material of heteromorph ammonites from the Lower Cretaceous of Lhoshag County, southern Tibet and its biostratigraphic implications. *Palaeoworld*, 33(3), 768-776. <https://doi.org/10.1016/j.palwor.2023.06.004>
- Maróti, B., *et al.* (2020). Joint application of structured-light optical scanning, neutron tomography and position-sensitive prompt gamma activation analysis for the non-destructive structural and compositional characterization of fossil echinoids. *NDT & E International*, 115. <https://doi.org/10.1016/j.ndteint.2020.102295>
- McCraw, J. R. C., Tobin, T. S., Cochran, J. K., & Landman, N. H. (2024). Ammonites as paleothermometers: Isotopically reconstructed temperatures of the Western Interior Seaway track global records. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 656. <https://doi.org/10.1016/j.palaeo.2024.112594>
- Moclán, A., *et al.* (2024). Deep learning identification of anthropogenic modifications on a carnivore remain suggests use of hyena pelts by Neanderthals in the Navalmaillo rock shelter (Pinilla del Valle, Spain). *Quaternary Science Reviews*, 329. <https://doi.org/10.1016/j.quascirev.2024.108560>
- Nathanail, A. (2023). Value of the data. *Data in Brief*, 48, 109188. <https://doi.org/10.5281/zenodo.7510741>

- Niño-García, A., Parra-Mosquera, J. D., & Macías-Villarraga, P. A. (2019). Upper Cretaceous chondrichthyes teeth record in phosphorites of the Loma Gorda Formation. *Boletín de Ciencias de la Tierra*, 46, 27-32. <https://doi.org/10.15446/rbct.n46.78525>
- Olsen, N., Chen, Y., Turberg, P., Moreau, A., & Alahi, A. (2025). X-ray Micro-CT based characterization of rock cuttings with deep learning. *Applied Computing and Geosciences*, 25. <https://doi.org/10.1016/j.acags.2025.100220>
- Pittman, M., Kaye, T. G., Graham, E., & Thorold, D. (2022). Laser-stimulated fluorescence in archaeology: Non-destructive fluorescence imaging for museum and field settings. *Journal of Archaeological Science: Reports*, 44. <https://doi.org/10.1016/j.jasrep.2022.103475>
- Szives, O., et al. (2024). Report on the 7th International Meeting of the IUGS Lower Cretaceous Ammonite Working Group, the Kilian Group (Warsaw, Poland, 21st August 2022): State of the art on the current standard ammonite zonation of the Western Tethyan Mediterranean Province. *Cretaceous Research*, 153. <https://doi.org/10.1016/j.cretres.2023.105716>
- Tian, Z., Wu, Y., Wei, Y., Zhao, Y., Pan, C., & Wang, Y. (2025). Fusion of near-infrared and Raman spectroscopy with machine learning strategies: Non-destructive rapid assessment of freshness and TVB-N value prediction in Pacific white shrimp (*Litopenaeus vannamei*). *Food Research International*, 214. <https://doi.org/10.1016/j.foodres.2025.116561>
- Trayler, R. B., Lopes, L. E., Holroyd, P. A., Kim, S. L., & Southon, J. R. (2024). Rapid, effective screening of tar seep fossils for radiocarbon and stable isotope analysis. *Quaternary Geochronology*, 85. <https://doi.org/10.1016/j.quageo.2024.101631>
- Vašiček, Z., Rabrenović, D., Skupien, P., & Radulović, B. (2023). Lower Berriasian ammonites from Dedina (Golubac Mountains, eastern Serbia) and their biostratigraphic implication. *Cretaceous Research*, 151. <https://doi.org/10.1016/j.cretres.2023.105623>
- Wang, B., et al. (2022). Understanding geological reports based on knowledge graphs using a deep learning approach. *Computers & Geosciences*, 168. <https://doi.org/10.1016/j.cageo.2022.105229>
- Yu, C., et al. (2024). Artificial intelligence in paleontology. *Earth-Science Reviews*, 252. <https://doi.org/10.1016/j.earscirev.2024.104765>