

## INTELIGENCIA ARTIFICIAL EN EL ARTE RUPESTRE: LÍMITES, POSIBILIDADES Y DESAFÍOS DE UNA EVOLUCIÓN INCIERTA

ARTIFICIAL INTELLIGENCE IN ROCK ART RESEARCH:  
LIMITS, POSSIBILITIES, AND CHALLENGES OF AN UNCERTAIN EVOLUTION

Hugo de Lara López

Universidad Nacional de Educación a Distancia (UNED)  
hdelara@invi.uned.es  
<https://orcid.org/0009-0004-0564-7222>

Recepción: 04/12/2025. Aceptación: 11/01/2026  
Publicación on-line: 12/01/2024

**RESUMEN:** El reciente desarrollo de la inteligencia artificial (IA) ha planteado nuevas posibilidades para la documentación y el análisis del arte rupestre. Este artículo examina críticamente el potencial y los límites del uso de modelos abiertos de *machine learning* (ML) para la detección automática de motivos pictóricos. Tras revisar el estado de la cuestión y explicar los principios del ML aplicados a la arqueología, se evalúa empíricamente un modelo abierto (Roboflow 3.0 Object Detection) en cuatro paneles rupestres de la península ibérica. Los resultados muestran un rendimiento aceptable solo en contextos bien conservados, mientras que disminuye notablemente ante pigmentos desvaídos o soportes irregulares. Se concluye que el ML puede apoyar la investigación, pero no sustituir la interpretación arqueológica, y que su avance depende de repositorios públicos estandarizados y anotaciones expertas coherentes.

**Palabras clave:** Arte prehistórico; *Machine learning*; Visión por computador; Detección automática; Modelos supervisados.

**ABSTRACT:** The recent development of artificial intelligence (AI) has opened new possibilities for the documentation and analysis of rock art. This article critically examines the potential and the limits of using open machine learning (ML) models for the automatic detection of pictorial motifs. After reviewing the state of the art and outlining the principles of ML as applied to archaeology, we empirically evaluate an open model (Roboflow 3.0 Object Detection) on four rock art panels from the Iberian Peninsula. The results show acceptable performance only in well-preserved contexts, while accuracy decreases notably in the presence of faded pigments or irregular rock surfaces. We conclude that ML can support archaeological research but cannot replace expert interpretation, and that progress in this field depends on standardized public repositories and coherent expert annotations.

**Keywords:** Prehistoric art; Machine learning; Computer vision; Automatic detection; Supervised models.

**Cómo citar este artículo / How to cite this article:** de Lara López, H. (2026). Inteligencia artificial en el arte rupestre: límites, posibilidades y desafíos de una evolución incierta. *Salduie*, 25.2: 1-14. [https://doi.org/10.26754/ojs\\_salduie/sald.2025212652](https://doi.org/10.26754/ojs_salduie/sald.2025212652)

## 1. INTRODUCCIÓN

La incorporación de técnicas de inteligencia artificial (IA) en arqueología ha experimentado un importante crecimiento en la última década, impulsada por el aumento de recursos digitales, la accesibilidad de modelos preentrenados y la consolidación de estándares de gestión de datos en humanidades digitales (Wilkinson *et al.* 2016; Fiorucci *et al.* 2020; Bellat *et al.* 2025). Aunque estos avances han transformado áreas como la prospección remota, la documentación 3D o el análisis de superficies patrimoniales, especialmente a partir de productos de teledetección y análisis derivados empleados en la investigación del arte rupestre (Jalandoni 2021), su integración metodológica en el estudio de las manifestaciones rupestres continúa siendo embrionaria. Ello se debe, entre otros factores, a la complejidad formal, tipológica, estilística y técnica de los motivos, a la enorme variabilidad de los soportes rocosos y a la limitada disponibilidad de conjuntos de imágenes suficientemente grandes y homogéneos como para entrenar modelos robustos (Kowlessar *et al.* 2021; Horn *et al.* 2022).

En este contexto, el *machine learning* (ML), entendido como un conjunto de métodos de IA que aprenden patrones a partir de datos sin necesidad de programar reglas explícitas (Goodfellow *et al.* 2016; Chollet 2021), ofrece oportunidades aún poco exploradas para automatizar tareas tradicionalmente dependientes de la observación experta. Estudios recientes han demostrado que, en condiciones controladas, los modelos de ML pueden identificar representaciones pictóricas, segmentar pigmentos o clasificar tipologías estilísticas con niveles apreciables de eficacia (Jalandoni *et al.* 2022; Suhaimi *et al.* 2023; Huo *et al.* 2025). En paralelo, investigaciones centradas en predicción espacial han empleado modelos supervisados para estimar áreas con alta probabilidad de albergar arte rupestre, abriendo vías prometedoras para la prospección a gran escala (Bonald *et al.* 2024). Estas contribuciones, si bien relevantes, dependen en gran medida de desarrollos internos no siempre accesibles y de *datasets* cuidadosamente seleccionados, lo que limita su reproducibilidad y adopción por parte de la comunidad arqueológica (Bevan 2015; Casillo *et al.* 2024).

Otro factor que ha frenado la expansión del ML en arte rupestre es la ausencia de repositorios públicos con imágenes anotadas y metadatos estandarizados. Mientras que otras áreas de la arqueología digital se han beneficiado de infraestructuras consolidadas,

siendo este el caso de la fotogrametría aplicada a patrimonio (Historic England 2017), los modelos 3D interoperables (Bruno *et al.* 2013; Dell'Unto *et al.* 2016) o los sistemas de documentación semántica basados en ontologías (Bruseker *et al.* 2017), el arte rupestre carece aún de conjuntos de datos abiertos, extensos y representativos que permitan entrenar modelos de forma transparente y comparativa. Este déficit afecta no solo al rendimiento de los modelos, sino también a la calidad de las evaluaciones, las cuales suelen realizarse sobre imágenes seleccionadas *ad hoc* y sin posibilidad de replicación independiente.

En este escenario, resulta imprescindible clarificar qué tipo de IA se está aplicando en cada caso. Con frecuencia, los estudios que emplean algoritmos clásicos de ML o variantes ligeras de redes neuronales profundas se presentan bajo el paraguas generalista de "inteligencia artificial", dificultando valorar el alcance real de cada técnica. La distinción no es menor: mientras que el *deep learning* (DL) requiere volúmenes masivos de datos y capacidades de cómputo elevadas (LeCun *et al.* 2015), los enfoques de ML más accesibles, incluyendo modelos preentrenados adaptables mediante *transfer learning* (Pan y Yang 2010; Zhuang *et al.* 2019), han demostrado su utilidad en contextos patrimoniales y artísticos con conjuntos de datos limitados (Sabatelli *et al.* 2018). Estos enfoques se ajustan mejor a los retos del arte rupestre, donde la escasez documental es estructural. Por ello, este trabajo se centra en métodos de ML que pueden ser utilizados por equipos arqueológicos sin infraestructuras avanzadas, respondiendo a la petición expresada en la bibliografía específica de explorar enfoques realistas, escalables y reproducibles (Sokolovska 2024).

En paralelo, comienzan a explorarse propuestas experimentales que emplean inteligencia artificial generativa para aumentar artificialmente conjuntos de datos patrimoniales mediante la creación de variaciones controladas de imágenes (Cioni *et al.* 2023). Pese a que estas aproximaciones no se han aplicado de forma directa al arte rupestre, sí apuntan a la posibilidad de simular condiciones variables de iluminación, deterioro o registro para complementar *datasets* reales.

En el ámbito más amplio del patrimonio digital, también se ha documentado el uso emergente de herramientas generativas en el diseño de recursos interactivos y visuales (Veggi 2025), confirmando una presencia creciente de estas técnicas en la investigación patrimonial. No obstante, su uso en arte rupestre plantea riesgos significativos, especialmente en un

contexto marcado por la proliferación de imágenes generadas sin control arqueológico. Por ello, su aplicación debe enmarcarse en protocolos estrictamente supervisados y nunca puede sustituir a la documentación real, sino únicamente ampliarla de manera experimental y controlada.

En definitiva, la aplicación del ML al arte rupestre se encuentra en un punto de inflexión, ya que existen avances significativos, pero también lagunas metodológicas, conceptuales y de infraestructura que dificultan valorar su alcance real.

Este trabajo se propone abordar estas cuestiones mediante tres objetivos complementarios: en primer lugar, presentar un marco actualizado y crítico de las investigaciones que han aplicado ML al arte rupestre y, en menor medida, al patrimonio arqueológico. En segundo lugar, explicar de manera accesible qué es el *machine learning*, cómo funciona y qué requisitos metodológicos exige su aplicación en contextos rupestres. En tercer lugar, evaluar de forma empírica un modelo abierto de detección de objetos entrenado parcialmente con imágenes de arte rupestre, con el objetivo de analizar sus posibilidades y limitaciones en diferentes escenarios de conservación y complejidad tipológica. Con ello, se pretende aportar una base sólida para debates futuros sobre el papel real (y no solo potencial) de la inteligencia artificial en la investigación del arte rupestre. Las conclusiones que se derivan de esta evaluación empírica se interpretan de forma conjunta con el marco teórico y bibliográfico revisado, siendo conscientes de que las pruebas realizadas tienen un carácter necesariamente limitado y exploratorio.

## 2. ESTADO DE LA CUESTIÓN

La aplicación del *machine learning* (ML) al estudio del arte rupestre constituye un campo aún emergente dentro de la arqueología digital. Aunque en los últimos años se ha producido un incremento de investigaciones, su desarrollo sigue condicionado por tres factores estructurales: la escasez de conjuntos de datos amplios y anotados, la diversidad formal, tipológica, estilística y técnica de las manifestaciones rupestres y la falta de herramientas abiertas que permitan la reproducibilidad metodológica (Bevan 2015; Fiorucci *et al.* 2020; Bellat *et al.* 2025). A pesar de estas limitaciones, comienzan a consolidarse varias líneas de trabajo que permiten vislumbrar tanto el alcance como las restricciones actuales de estas tecnologías.

### 2.1. Detección y clasificación automática de motivos

La primera línea de investigación se centra en la identificación y clasificación automática de figuras pictóricas a partir de documentación fotográfica. Este enfoque ha empleado principalmente modelos de detección de objetos basados en arquitecturas de *deep learning* como YOLO o Faster/Mask R-CNN, adaptadas mediante *transfer learning* para compensar conjuntos de datos reducidos (Redmon *et al.* 2016; He *et al.* 2017; Pan y Yang 2010). Algunos de los primeros intentos de aplicar *machine learning* al arte rupestre ya exploraron la detección automatizada de motivos mediante enfoques tempranos de ML apoyados en infraestructuras en la nube, anticipando problemas de escalabilidad y acceso remoto que siguen siendo relevantes hoy (Ponmagal y Srinivasan 2015).

Uno de los trabajos pioneros es el de Jalandoni, Zhang y Zaidi (2022), quienes demostraron que modelos convolucionales pueden reconocer motivos en contextos de variabilidad lumínica y fondos complejos, alcanzando niveles de precisión superiores a los métodos manuales en contextos controlados. Por su parte, Suhaimi *et al.* (2023) compararon estrategias de una y dos etapas (YOLO frente a Faster R-CNN), mostrando que los modelos de una etapa resultan más eficientes en escenarios con recursos computacionales limitados, mientras que las arquitecturas más profundas logran mayor precisión en paneles con pigmentos desvaídos.

En un enfoque similar, Kowlessar *et al.* (2021) aplicaron *transfer learning* para clasificar estilos en Arnhem Land (Australia), demostrando que incluso con *datasets* pequeños es posible obtener resultados robustos cuando las clases están bien definidas y la documentación es homogénea. En una línea complementaria, Turner-Jones *et al.* (2024) exploraron la clasificación de motivos rupestres mediante ML en un entorno de ciencia ciudadana con fines educativos, subrayando el papel central de la anotación humana y la supervisión experta incluso en aplicaciones asistidas por IA.

### 2.2. Segmentación de pigmentos y reconstrucción de contornos

La segmentación, tanto semántica como por instancia, representa otro avance relevante en el estudio del arte rupestre. La implementación reciente de modelos

como YOLOv7-Seg en contextos rupestres (Huo *et al.* 2025) ha permitido generar máscaras píxel a píxel que reconstruyen contornos incluso cuando los pigmentos presentan deterioro moderado. Estos modelos funcionan especialmente bien cuando las figuras mantienen continuidad formal, pero experimentan descensos de rendimiento ante soportes fracturados, superposiciones o pigmentos muy desvaídos. Asimismo, se han desarrollado enfoques no supervisados y semisupervisados para agrupar estilos y detectar patrones sin necesidad de grandes volúmenes de etiquetas (Schmarje *et al.* 2020; Yang *et al.* 2021; Chen *et al.* 2024). Estas técnicas resultan útiles para detectar similitudes formales; sin embargo, su capacidad interpretativa depende de la coherencia de las muestras y del diseño de las características empleadas.

### 2.3. Predicción espacial y prospección asistida por IA

Paralelamente, han surgido investigaciones centradas en la predicción de áreas con alta probabilidad de albergar arte rupestre. Bonald *et al.* (2024) utilizaron modelos de ML para identificar zonas potenciales en la cuenca del Pajeú (Brasil), combinando variables ambientales y geográficas. Este tipo de estudios se sitúa en una línea intermedia entre la teledetección arqueológica clásica y los modelos predictivos automatizados, mostrando la capacidad de la IA para ampliar la escala de prospección más allá de las áreas tradicionalmente estudiadas.

En esta línea deben mencionarse también los trabajos de teledetección y análisis de superficies mediante fotogrametría, LiDAR o imágenes multispectrales, los cuales constituyen una base imprescindible para cualquier *pipeline* de ML (Bruno *et al.* 2013; Historic England 2017). Aunque estos métodos no siempre se desarrollan específicamente en clave de ML, proporcionan los datos fundamentales sobre los que se entrenan y evalúan los modelos de detección.

### 2.4. Gestión patrimonial, documentación 3D y análisis de superficies

Al margen de los estudios estrictamente centrados en arte rupestre, la literatura arqueológica ha producido avances relevantes aplicables indirectamente al campo. Fiorucci *et al.* (2020) realizaron una revisión

exhaustiva del ML en conservación patrimonial, destacando su uso en clasificación de materiales, análisis de degradaciones y predicción de riesgos. Por su parte, Liu *et al.* (2023) exploraron la detección de anomalías en superficies pétreas mediante autoencoders y GANs, técnicas altamente pertinentes para la identificación de alteraciones que pueden confundirse con trazos pictóricos. En su caso, Casillo *et al.* (2024) evaluaron el potencial de la IA en la gestión de los yacimientos, subrayando una necesidad de marcos metodológicos reproducibles y de datos FAIR.

En una línea metodológicamente afín, Monna *et al.* (2020) demostraron la eficacia del *machine learning* para el mapeo rápido de estructuras arqueológicas en piedra seca, evidenciando tanto el potencial de estas técnicas en contextos visualmente complejos como su dependencia de una documentación previa homogénea y bien controlada. En este marco debemos incluir también los enfoques computacionales avanzados de segmentación interactiva en 3D aplicados específicamente al arte rupestre, los cuales combinan mapas de profundidad y técnicas de regularización de gradientes para mejorar la delimitación de figuras sobre soportes complejos (Zeppelzauer *et al.* 2016).

Al mismo tiempo, estudios como los de Dell'Unto *et al.* (2016) o Bruno *et al.* (2013) han consolidado procedimientos de documentación 3D que facilitan la creación de modelos digitales completos, esenciales tanto para el entrenamiento de modelos como para la validación arqueológica de los resultados.

### 2.5. Enfoques generativos y ampliación sintética de datos

Una tendencia reciente es la aplicación de IA generativa para aumentar artificialmente los conjuntos de datos disponibles. Cioni *et al.* (2023) demostraron cómo los modelos de difusión pueden generar variaciones sintéticas útiles para tareas de *captioning* y *retrieval* en contextos patrimoniales, mientras que Veggi (2025) revisó su potencial en el diseño de experiencias interactivas. Aunque estas técnicas podrían ayudar a mitigar la escasez documental del arte rupestre, su uso requiere extrema cautela debido a la proliferación de imágenes falsas y a los riesgos epistemológicos asociados.

En un uso estrictamente controlado, podrían emplearse de forma experimental para generar variaciones sintéticas a partir de las imágenes reales

documentadas, simulando, por ejemplo, cambios de iluminación, contraste o grados progresivos de deterioro, con el único objetivo de reforzar el entrenamiento preliminar de modelos de detección y evaluar su robustez frente a condiciones visuales habituales en el trabajo de campo.

No obstante, la incorporación no supervisada de imágenes generadas puede contaminar los conjuntos de datos, reforzar sesgos formales o dificultar la distinción entre documentación empírica y material sintético. Por ello, su aplicación en arte rupestre debe quedar limitada a entornos experimentales con trazabilidad completa de los datos y supervisión arqueológica directa.

## 2.6. Limitaciones estructurales y desafíos actuales

La literatura coincide en señalar varios problemas transversales:

- Escasez y heterogeneidad de datos: la falta de repositorios públicos anotados impide entrenar modelos sólidos y reproducibles (Bevan 2015; Fiorucci *et al.* 2020).
- Diversidad formal, tipológica, estilística y técnica del arte rupestre: no existe un conjunto unificado de categorías aplicable a todas las regiones. Comparar tipologías, estilos y técnicas distintas resulta metodológicamente problemático (Sokolovska 2024).
- Dependencia de desarrollos internos: muchos avances provienen de los centros con acceso a recursos computacionales u hojas de datos que no son accesibles, lo cual limita su adopción generalizada (Horn *et al.* 2022; Casillo *et al.* 2024).
- Falta de protocolos comunes: las guías de anotación, los criterios de segmentación y las métricas varían notablemente entre los distintos estudios, dificultando las comparaciones sistemáticas.

En conjunto, el estado de la cuestión muestra un campo en rápida expansión, pero metodológicamente desigual, donde coexisten avances prometedores y limitaciones profundas. La literatura apunta a la necesidad urgente de repositorios públicos, marcos metodológicos comunes y evaluaciones contextualizadas que permitan valorar con precisión el alcance real del machine learning en el estudio del arte rupestre.

## 3. ¿QUÉ ES EL MACHINE LEARNING Y CÓMO SE APLICA AL ARTE RUPESTRE?

El *machine learning* (ML) constituye un conjunto de métodos de la inteligencia artificial que permiten a los sistemas informáticos aprender patrones a partir de ejemplos, sin necesidad de programar reglas explícitas para cada tarea (Goodfellow *et al.* 2016; Chollet 2021). En términos generales, un modelo ajusta sus parámetros internos (normalmente millones de valores distribuidos en capas neuronales) para transformar una entrada (como la imagen de un panel rupestre) en una salida útil, ya sea localizar un motivo, clasificarlo o segmentarlo. Este proceso se articula mediante el entrenamiento: el modelo compara sus predicciones con las anotaciones proporcionadas por especialistas, calcula su error mediante una función de pérdida y modifica sus parámetros para reducirlo progresivamente (LeCun *et al.* 2015).

En el ámbito del arte rupestre, estas capacidades permiten abordar distintos tipos de problemas. El aprendizaje supervisado se emplea para detectar motivos, clasificarlos tipológicamente o segmentar su contorno, incluso en condiciones de alteración avanzada (Suhaimi *et al.* 2023; Huo *et al.* 2025). El aprendizaje no supervisado ofrece herramientas para agrupar imágenes por semejanza sin etiquetas previas, lo que puede ayudar a explorar similitudes formales, tipológicas, estilísticas o técnicas (Chen *et al.* 2024). Entre ambos enfoques, los métodos semi-supervisados aprovechan pequeños conjuntos anotados junto a grandes volúmenes de imágenes sin etiquetar, una estrategia especialmente valiosa en disciplinas como la arqueología, donde la documentación disponible es limitada (Schmarje *et al.* 2020; Yang *et al.* 2021).

Para aplicar ML a problemas arqueológicos es necesario traducir la pregunta científica en un problema computacional bien formulado. Este paso, que en la literatura técnica se denomina *problem formulation* más que “parametrización”, implica definir con claridad qué se desea obtener: detecciones, segmentaciones, clases estilísticas, mapas de probabilidad o predicción espacial. Cada tipo de tarea requiere estructurar los datos de forma diferente, ya sea mediante etiquetas globales, cajas delimitadoras (*bounding boxes*), máscaras a nivel de píxel, nubes de puntos o modelos de reflectancia (Bruno *et al.* 2013; Historic England 2017).

Las elecciones metodológicas posteriores (tipo de arquitectura, función de pérdida, hiperparámetros)



deben ser coherentes con esta formulación. La función de pérdida define cómo se mide el error entre la predicción y la anotación real. No siempre corresponde a una noción estricta de “distancia”, pues algunas funciones son probabilísticas, otras combinan varios términos (clasificación, localización, segmentación). Los hiperparámetros, por su parte, son ajustes externos al modelo que controlan su proceso de aprendizaje: tasa de aprendizaje, tamaño del lote (batch size), número de iteraciones, técnicas de regularización o estrategias de aumento de datos. La arquitectura (número de capas, tipo de convoluciones) no se “parametriza” durante el entrenamiento, sino que se diseña previamente.

El éxito del ML depende de la calidad del conjunto de datos. En arte rupestre, donde predominan imágenes heterogéneas en iluminación, color, resolución y estado de conservación, resulta esencial documentar adecuadamente el contexto (Wilkinson *et al.* 2016; Bruseker *et al.* 2017). Las anotaciones deben ser consistentes y reproducibles entre especialistas, especialmente en casos con pigmentos fragmentados o superposiciones. Asimismo, para evitar la sobreestimación del rendimiento, es necesario separar estrictamente las imágenes de entrenamiento y prueba por paneles o yacimientos, como recomienda Bevan (2015).

En cuanto a los modelos, se utilizan habitualmente arquitecturas de detección y segmentación de última generación, como la familia YOLO o Mask R-CNN (Redmon *et al.* 2016; He *et al.* 2017). Su entrenamiento se basa con frecuencia en el *transfer learning*, el cual permite adaptar modelos previamente entrenados en grandes bases de datos generales a tareas arqueológicas más específicas (Pan y Yang 2010; Zhuang *et al.* 2019). Las técnicas de aumento de datos (rotaciones, recortes, variaciones de iluminación) (Shorten y Khoshgoftaar 2019) ayudan a compensar la escasez de documentación típica de este campo.

La aplicación de un modelo en fase de inferencia requiere interpretar sus resultados con cautela. Los umbrales de confianza y los algoritmos de supresión de solapamientos (*Non-Maximum Suppression*, NMS) influyen directamente en la aparición de falsos positivos y falsos negativos (Guo *et al.* 2017). En este sentido, las métricas deben contextualizarse, puesto que maximizar la exhaustividad puede ser adecuado para la prospección de nuevos paneles, mientras que en catalogaciones definitivas se prioriza la precisión (Bellat *et al.* 2025). Desde una perspectiva arqueológica, estas métricas no son neutras, ya que los falsos posi-

tivos y los falsos negativos tienen implicaciones distintas según los objetivos de la investigación. En contextos de prospección o cribado preliminar, puede resultar preferible asumir un mayor número de falsos positivos a cambio de maximizar la detección potencial de motivos, aun a costa de una posterior revisión manual. Por el contrario, en tareas de catalogación sistemática o documentación exhaustiva, los falsos positivos adquieren un mayor peso, al introducir elementos espurios que afectan a la calidad del registro y a su interpretación posterior.

La evaluación del rendimiento de los modelos debe, por tanto, interpretarse siempre en relación con el propósito arqueológico concreto para el que se aplican y no únicamente a partir de valores métricos absolutos importados del ámbito de la visión por computador. El examen de los errores (confusión con grietas, líquenes, sombras, manchas de agua o pérdidas de pigmento) nos ofrece información valiosa para mejorar la anotación y la formulación del problema.

Finalmente, la aplicación del ML al arte rupestre exige considerar aspectos éticos y de reproducibilidad como documentar los datos utilizados, versionar los conjuntos de imágenes y describir claramente las limitaciones y sesgos del modelo (Dell'Unto *et al.* 2016; Wilkinson *et al.* 2016). Lejos de reemplazar la interpretación arqueológica, estos sistemas actúan como herramientas de apoyo que amplían la capacidad de observación y aceleran la revisión de grandes volúmenes de datos, siempre en estrecho diálogo con la mirada experta del arqueólogo.

#### 4. EVALUACIÓN PRÁCTICA DE UN MODELO ABIERTO DE DETECCIÓN DE OBJETOS EN ARTE RUPESTRE

Con el fin de aproximarnos de manera empírica a las posibilidades y limitaciones del *machine learning* en el estudio del arte rupestre, se ha seleccionado el modelo *Roboflow 3.0 Object Detection (Fast)*, entrenado a partir de un *checkpoint* preentrenado en COCO y publicado en Roboflow Universe bajo la denominación *RMR Computer Vision Model*, perteneciente al usuario Manifestaciones y figuras rupestres. Se trata de un modelo de detección de objetos que está basado en *deep learning*, optimizado para procesar imágenes de forma rápida y generar predicciones que incluyen la localización y la clase de los elementos identificados.

La elección de este modelo responde a tres criterios clave derivados de la literatura reciente: en primer lugar, su condición de recurso abierto y reproducible, en contraste con los desarrollos internos no accesibles que caracterizan a gran parte de las investigaciones actuales (Horn *et al.*, 2022; Casillo *et al.*, 2024). En segundo lugar, la disponibilidad pública de sus métricas y datos de entrenamiento, que es un requisito esencial para una evaluación transparente (Wilkinson *et al.*, 2016). En tercer lugar, su relativa estabilidad, evidenciada en la consistencia de sus valores de validación.

Estas pruebas no pretenden ofrecer una validación exhaustiva del modelo, sino explorar su comportamiento en contextos rupestres reales y contrastar sus resultados con las limitaciones señaladas por la bibliografía especializada. Conviene precisar, no obstante, que el modelo evaluado se inscribe plenamente en el paradigma del *deep learning*, pese a su carácter abierto y accesible. Al igual que ocurre con la mayoría de las herramientas actualmente disponibles para la comunidad investigadora, su funcionamiento depende de arquitecturas neuronales profundas y de preentrenamientos masivos sobre grandes conjuntos de datos generales como COCO. Este hecho no contradice el enfoque defendido en este trabajo, sino que, al contrario, refuerza su planteamiento crítico, pues incluso los modelos considerados “ligeros”, reutilizables o de fácil acceso continúan apoyándose en infraestructuras conceptuales y técnicas propias del *deep learning*. La evaluación de este tipo de herramientas permite, precisamente, poner de manifiesto que muchas de las limitaciones asociadas a estas arquitecturas (dependencia de grandes volúmenes de datos, sesgos del preentrenamiento o dificultades para generalizar en contextos visuales complejos) persisten también en entornos abiertos y reproducibles.

El modelo fue entrenado originalmente sobre un conjunto de categorías arqueológicas muy heterogéneo (amolador, batea-metate, cúpula, dolmen, esfera lítica, geoglifo, menhir, micropetroglifo, monolito, petroglifo y pintura rupestre). Esta diversidad tiene implicaciones metodológicas relevantes. Por un lado, explica la amplitud conceptual de la categoría “pintura rupestre”, que funciona en el modelo como una clase contenedora que agrupa todo el repertorio pictórico del *dataset*. Por otro lado, el modelo ayuda a comprender algunas de las confusiones observadas en las pruebas empíricas, ya que el sistema ha aprendido patrones visuales distribuidos entre dominios muy distintos.

El conjunto total de datos empleados en el entrenamiento asciende a 7.679 imágenes, organizadas de la siguiente manera:

- *Train set*: 6.717 imágenes (87 %), utilizadas para el aprendizaje de los patrones visuales.
- *Valid set*: 640 imágenes (8 %), destinadas al ajuste intermedio de hiperparámetros y a la evaluación durante el proceso de entrenamiento. En este subconjunto, la categoría “pintura rupestre” obtuvo un mAP@50 del 45 %.
- *Test set*: 322 imágenes (4 %), reservadas para medir la capacidad de generalización sobre imágenes nunca vistas, con un mAP@50 del 54 % en dicha categoría.

Todas las imágenes fueron redimensionadas mediante *stretching* (no proporcional) a 640 × 640 píxeles. Este preprocesamiento favorece el rendimiento computacional, pero puede introducir pequeñas distorsiones que afectan especialmente a paneles con soportes muy irregulares, un aspecto señalado repetidamente en la literatura arqueológica como un desafío para la interpretación automatizada (Fiorucci *et al.*, 2020). Con el objetivo de mejorar la robustez del modelo frente a las condiciones variables del registro rupestre, se aplicaron técnicas de aumento de datos que generaron tres variantes aumentadas por imagen, en función de los parámetros establecidos. Estas transformaciones incluyeron: rotaciones aleatorias entre -15° y +15°, conversión a escala de grises en un 25 % de los casos, modificaciones del tono (±15 %), variaciones de saturación (-25 % a +25 %) y ajustes de brillo (±15 %). Estos aumentos buscan simular situaciones habituales en el trabajo de campo (diferencias de iluminación, contraste, ángulos de captura o pérdida parcial de pigmentos) y se ajustan a las recomendaciones metodológicas establecidas para paliar la escasez de datos arqueológicos (Shorten y Khoshgoftaar 2019).

La evolución del entrenamiento puede observarse en las gráficas del proceso de entrenamiento y validación (Fig. 1). En ellas se aprecia una reducción progresiva de la pérdida (*loss*) tanto en el conjunto de entrenamiento como en el de validación, junto con la estabilización de las métricas de *precision*, *recall* y mAP. Este comportamiento indica un aprendizaje consistente y una capacidad razonable de generalización hacia imágenes no incluidas en el entrenamiento, coherente con lo esperado en modelos basados en arquitecturas modernas de detección (Goodfellow *et al.*, 2016). La convergencia paralela de ambas curvas sugiere, además, la ausencia de sobreajuste severo,

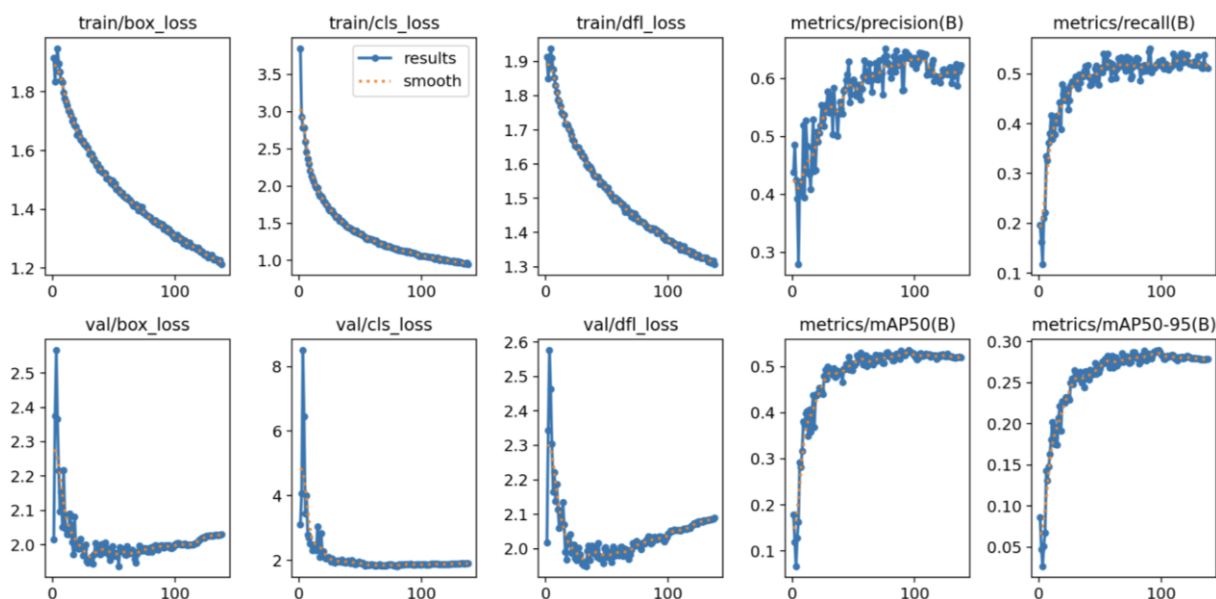


Figura. 1. Gráficas del proceso de entrenamiento y validación del modelo Roboflow 3.0 Object Detection (Fast). Se observa la reducción progresiva de los errores (loss) y la estabilización de las métricas de precisión, recall y mAP, lo cual parece indicar un aprendizaje sólido del sistema en la identificación de elementos arqueológicos.

un aspecto especialmente relevante dada la heterogeneidad de los repertorios rupestres incluidos en el conjunto de datos.

El rendimiento final del modelo se evaluó mediante métricas estandarizadas en visión por computador: un mAP@50 global del 52,8 %, una *precision* del 62,4 % y un *recall* del 52,4 %. A pesar de que estos valores muestran una capacidad apreciable para identificar pinturas rupestres en contextos visualmente variados, la presencia de falsos positivos y negativos confirma que los resultados deben ser interpretados de manera crítica y en diálogo con la lectura experta.

La categoría “pintura rupestre”, única destinada a recoger todos los motivos pictóricos del *dataset*, agrupa realidades formales, tipológicas, estilísticas y técnicas muy diversas, lo cual limita inevitablemente la capacidad discriminativa del modelo y refleja una de las principales carencias que son señaladas por la crítica académica: la insuficiencia de taxonomías detalladas y anotaciones consistentes (Sokolovska, 2024).

Desde una perspectiva arqueológica, los resultados indican que los modelos abiertos como este pueden servir como herramientas de apoyo para la revisión inicial de grandes conjuntos de imágenes, la detección preliminar de áreas de interés o la identificación de figuras bien preservadas. No obstante, sus limitaciones se vuelven evidentes en contextos com-

plejos con superposiciones, pérdida de pigmentos o soportes alterados, tal como se analizará en el apartado siguiente. La aplicación de ML al arte rupestre demanda, por tanto, una combinación equilibrada entre automatización y control experto, así como el desarrollo de conjuntos de datos más amplios, anotados bajo criterios homogéneos y accesibles públicamente.

Para las pruebas empíricas se seleccionaron imágenes procedentes de repositorios institucionales que garantizan una calidad visual adecuada para la evaluación automática. La única excepción es el panel 4 de la Cueva de Atlanterra, documentado durante las campañas de campo dirigidas por Martí Mas en la década de los años 90 (De Lara, 2025; De Lara *et al.*, 2025).

#### 4.1. Evaluación del modelo en los contextos rupestres ibéricos: pruebas y análisis crítico

Para evaluar la aplicabilidad real del modelo en contextos arqueológicos y no únicamente desde parámetros métricos, se sometió el sistema a un conjunto de pruebas sobre paneles rupestres ubicados en la península ibérica que representan distintos grados de conservación y complejidad visual. La selección de los casos responde al objetivo de cubrir un espectro representativo de situaciones habituales en el estudio



del arte rupestre, por lo que hemos considerado motivos nítidos y completos, figuras fragmentadas por alteraciones del soporte y paneles con alta densidad de elementos registrados en condiciones lumínicas adversas.

La primera prueba se llevó a cabo con un panel del Abrigo de Peña Escrita (Fuencaliente, Ciudad Real) (Figs. 2.1a y 2.1b), el cual está bien conservado, pero presenta tipologías menos frecuentes dentro del repertorio documentado en el *dataset* de entrenamiento. Este caso resulta especialmente útil, ya que permite valorar hasta qué punto el modelo es capaz de reconocer figuras que se apartan de los patrones más frecuentes presentes en su entrenamiento. El sistema detectó algunas de las figuras con umbrales de confianza que alcanzan o superan puntualmente el 50 %, si bien la mayoría quedaron por debajo de ese valor. La omisión de detalles pequeños y ciertas identificaciones incorrectas sugieren que el modelo experimenta dificultades cuando afronta motivos que no se ajustan a las morfologías predominantes de su conjunto de entrenamiento, confirmando la fuerte dependencia de estos modelos respecto a distribuciones de datos homogéneas.

La segunda prueba se aplicó a un panel de los Abrigos de Peña del Escrito (Villar del Humo, Cuenca) (Figs. 2.2a y 2.2b), donde las figuras presentan cierto grado de alteración. En este caso, el modelo consiguió identificar los dos motivos principales, aunque con umbrales de confianza marcadamente desiguales. Si bien la capacidad para reconocer figuras completas resulta positiva, la interpretación errónea de cortes naturales del soporte como un geoglifo refleja una de las limitaciones más señaladas en la literatura: la confusión entre trazos intencionales y alteraciones naturales cuando los pigmentos están fragmentados o desvanecidos.

La tercera prueba se desarrolló en el panel de Solana de las Covachas (Nerpio, Albacete) (Figs. 2.3a y 2.3b), caracterizado por una fuerte alteración que divide los pigmentos en segmentos inconexos. En este contexto, el modelo no logró identificar el motivo principal como una unidad, generando delimitaciones imprecisas y fragmentando la figura en detecciones parciales. Este comportamiento evidencia una limitación inherente a los modelos actuales: su dificultad para reconstruir la continuidad formal de un motivo cuando el pigmento ha perdido cohesión visual. El arqueólogo, en cambio, es capaz de inferir esa unidad a partir de criterios que no están presentes en los datos de entrenamiento.

Por último, se evaluó el modelo en condiciones especialmente adversas mediante una imagen del panel 4 de la Cueva de Atlanterra (Tarifa, Cádiz) (Figs. 2.4a y 2.4b), caracterizada por escasa iluminación, alta densidad de motivos, múltiples superposiciones y pigmentos desvaídos. En este escenario, el rendimiento del sistema fue claramente insuficiente, pues gran parte de los motivos no fueron detectados, las delimitaciones resultaron muy imprecisas y varias estructuras naturales fueron confundidas con figuras. Este desempeño era previsible, ya que la escena se sitúa fuera del rango de variabilidad presente en el *dataset* de entrenamiento, confirmando que estos modelos funcionan de manera aceptable en contextos simples o moderados, pero fallan cuando se enfrentan a contextos complejos.

En conjunto, estas pruebas señalan hacia una tendencia coherente con la literatura especializada: los modelos abiertos basados en *deep learning* ofrecen resultados razonables en figuras bien conservadas y tipologías frecuentes, pero su rendimiento flaquea ante soportes alterados, estructuras complejas o pigmentos desvanecidos. Los errores recurrentes (fragmentación de motivos, confusión con grietas y omisión de pigmentos desvaídos) confirman que la visión automática debe entenderse hoy como una herramienta auxiliar y no sustitutiva de la interpretación experta. Estas pruebas permiten, además, aproximar-nos a un marco realista sobre lo que puede y no puede ofrecer el *machine learning* en el estudio del arte rupestre en su estado actual.

#### 4.2. Discusión: implicaciones arqueológicas y metodológicas

Los resultados obtenidos permiten plantear una serie de reflexiones que trascienden la evaluación concreta del modelo y se relacionan con los desafíos más amplios señalados por la literatura especializada. En primer lugar, las pruebas confirman que el rendimiento de los modelos de detección depende estrechamente de la continuidad formal y cromática de los motivos. Allí donde las figuras presentan contornos bien definidos, pigmentos relativamente homogéneos o morfologías claras, la detección es razonable, aunque con umbrales de confianza variables. Sin embargo, en contextos en los que la alteración del soporte interrumpe la morfología original, los modelos tienden a fragmentar los motivos o a confundir elementos naturales con representaciones pictóricas.

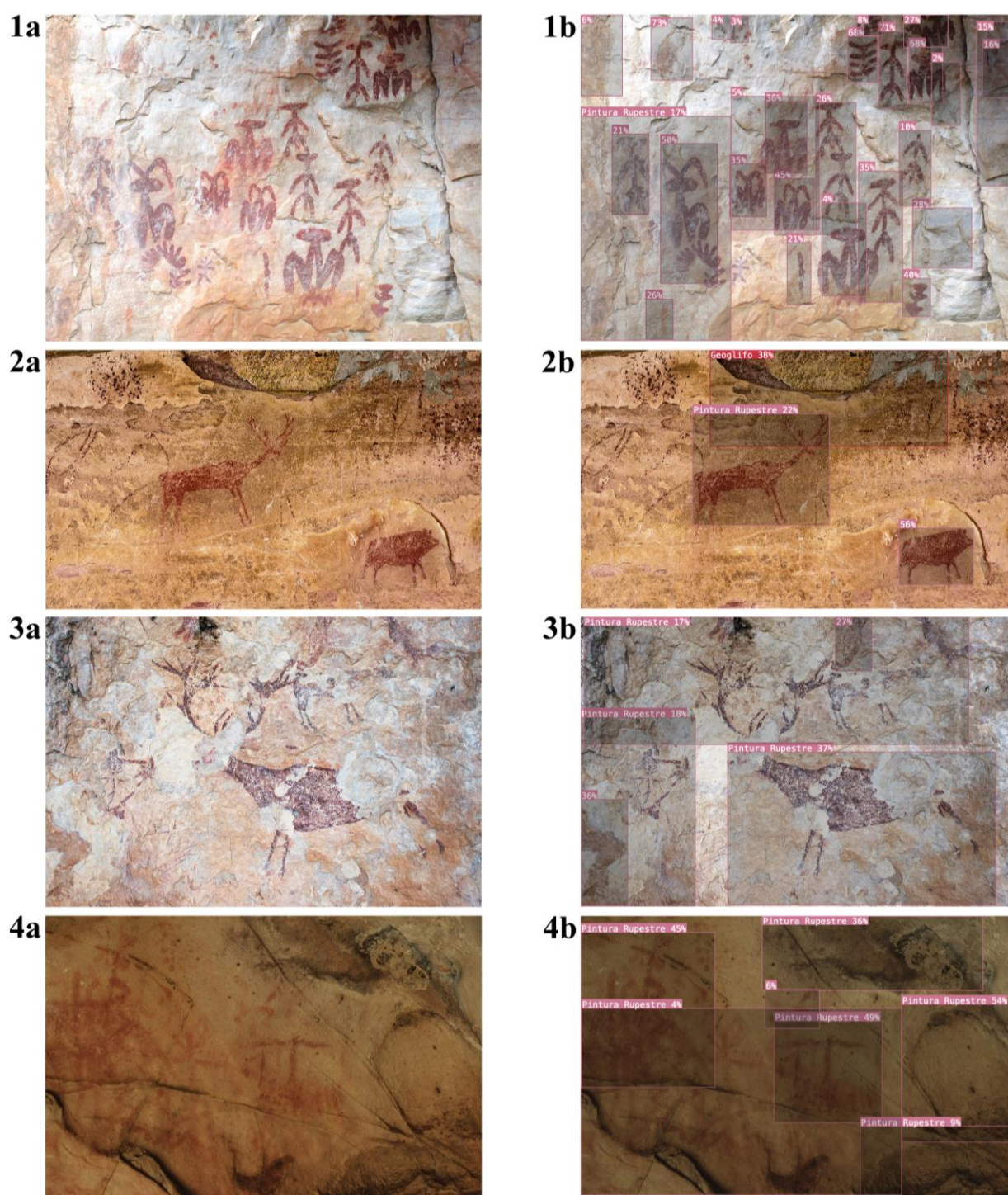


Figura 2. Evaluación del modelo en distintos contextos rupestres.

- 1a. Fotografía de uno de los paneles del Abrigo de Peña Escrita (repositorio institucional de la Consejería de Educación, Cultura y Deportes de Castilla-La Mancha).
- 1b. Resultados del modelo aplicado.
- 2a. Fotografía de un panel de los Abrigos de Peña del Escrito (repositorio institucional de la Consejería de Educación, Cultura y Deportes de Castilla-La Mancha).
- 2b. Resultados del modelo aplicado.
- 3a. Fotografía de un panel de Solana de las Covachas (repositorio institucional de la Consejería de Educación, Cultura y Deportes de Castilla-La Mancha).
- 3b. Resultados del modelo aplicado.
- 4a. Fotografía del sector derecho del panel 4 de la Cueva de Atlanterra (De Lara, 2025; De Lara et al., 2025).
- 4b. Resultados del modelo aplicado.



Esta dinámica mencionada refleja un problema estructural: las redes neuronales trabajan sobre correlaciones visuales locales, mientras que la interpretación arqueológica integra información contextual, formal, tipológica, estilística y técnica que no se encuentra explícitamente en las imágenes. El arqueólogo puede llegar a reconstruir mentalmente formas alteradas, identificar superposiciones o distinguir incluso una línea intencional de una grieta; el modelo, en cambio, carece de referencias externas que le permitan corregir ambigüedades o inferir continuidad allí donde el pigmento ha desaparecido. Este desfase evidencia que la visión automática puede asistir a la observación, pero no sustituir el razonamiento arqueológico. A todo ello se suma la cuestión de la representatividad de los datos.

Las dificultades del modelo en Solana de las Covachas y, especialmente, en Atlanterra, muestran que los conjuntos de entrenamiento actuales no abarcan la diversidad visual del arte rupestre peninsular, una limitación que también ha sido señalado por otros trabajos recientes sobre IA aplicada al patrimonio rupestre (Bevan 2015; Horn *et al.* 2022; Sokolovska 2024).

La falta de pigmentos continuos, las superposiciones, los paneles muy poblados y las condiciones lumínicas adversas quedan fuera del rango para el que estos modelos fueron entrenados, generando un rendimiento desigual que reproduce (y en ocasiones también amplifica) los sesgos del *dataset* (Suhaimi *et al.* 2023; Bellat *et al.* 2025; Huo *et al.* 2025). Este aspecto coincide con críticas recientes que subrayan la necesidad de construir repositorios amplios, abiertos y estandarizados, capaces de capturar la heterogeneidad del patrimonio rupestre más allá de recoger conjuntos emblemáticos o formalmente sencillos (Wilkinson *et al.* 2016; Bruseker *et al.* 2017; Casillo *et al.* 2024).

También resulta significativo que el modelo produzca errores recurrentes en la interpretación de las grietas, desconchados o sombras como figuras. Este fenómeno pone de relieve la importancia que tienen las estrategias de anotación y del diseño mismo del problema. La categoría "pintura rupestre", al englobar una variabilidad, se convierte en una etiqueta que resulta demasiado amplia y poco informativa desde un punto de vista arqueológico. Dividir esta categoría en clases más específicas o incorporar anotaciones negativas (regiones explícitamente marcadas como no pictóricas) podría mejorar la discriminación del modelo.

Por último, estas pruebas evidencian la necesidad de desarrollar protocolos metodológicos compartidos para la evaluación de modelos aplicados a patrimonio. Mientras que en otras áreas del *machine learning* los procedimientos de validación están consolidados, en arte rupestre persisten diferencias importantes en cómo se anotan los motivos, cómo se estructuran los conjuntos de entrenamiento y qué métricas se consideran relevantes. Nuestros resultados refuerzan la idea de que la reproducibilidad (tanto en la documentación como en la evaluación y en la liberación transparente de los modelos y sus parámetros entrenados) es una condición imprescindible para que la IA se integre plenamente en las prácticas de investigación arqueológica.

En conjunto, la discusión apunta hacia un marco en el cual el *machine learning* no reemplaza el análisis arqueológico, sino que lo complementa, ofreciendo capacidades útiles para agilizar tareas repetitivas y explorar grandes volúmenes de imágenes, pero siempre bajo supervisión experta. La clave para avanzar reside en mejorar los datos, refinar las categorías analíticas y, sobre todo, comprender qué tipo de preguntas arqueológicas pueden beneficiarse realmente de la automatización.

## 5. CONCLUSIONES

La aplicación del *machine learning* al estudio del arte rupestre se encuentra en estos momentos en una fase de desarrollo significativa pero todavía irregular, condicionada por la limitada disponibilidad de datos, la heterogeneidad de los soportes y la complejidad formal, tipológica, estilística y técnica de las manifestaciones rupestres.

Aunque las pruebas empíricas realizadas en este trabajo son necesariamente limitadas y no pretenden agotar la diversidad del registro rupestre peninsular, sus resultados resultan coherentes con los problemas estructurales identificados de forma reiterada por la bibliografía especializada. En este sentido, las conclusiones de alcance general que se plantean no se derivan exclusivamente del experimento, sino de su lectura conjunta con el marco teórico y los estudios previos revisados. Desde esta perspectiva, el valor de este trabajo no reside en la generalización directa de sus resultados empíricos, sino en mostrar, mediante un caso de estudio reproducible, cómo las limitaciones teóricas señaladas por la literatura se manifiestan de forma concreta en contextos rupestres reales.

En primer lugar, las pruebas confirman que los modelos de detección actuales pueden ofrecer un rendimiento aceptable en contextos que se encuentren bien conservados. En estas condiciones, la IA es capaz de localizar figuras y delimitar sus contornos con una precisión razonable, reduciendo el tiempo requerido para revisar grandes conjuntos de imágenes. Sin embargo, este rendimiento se degrada cuando los modelos se enfrentan a pigmentos fragmentados, superposiciones complejas, soportes irregulares o condiciones lumínicas adversas.

Las pruebas realizadas en Solana de las Cova-chas y Atlanterra ilustran claramente esta limitación: los errores se multiplican y adoptan patrones recurrentes (fragmentación de motivos, detección errónea de irregularidades naturales, omisión de pigmentos desvaídos o delimitaciones desplazadas) que evidencian la ausencia en los datos de entrenamiento de los criterios arqueológicos esenciales para su comprensión (continuidad gráfica, coherencia estilística, articulación espacial o identificación de las superposiciones).

En segundo lugar, los resultados corroboran uno de los desafíos más señalados por la bibliografía: la falta de repositorios amplios, representativos y estandarizados para entrenar modelos aplicados al patrimonio rupestre. La enorme variabilidad contextual, formal, tipológica, técnica y estilística del arte rupestre peninsular no puede ser abordada eficazmente sin conjuntos de datos extensos, consistentes y anotados bajo criterios arqueológicos homogéneos. Sin estos cimientos, cualquier avance en el campo seguirá reproduciendo (e incluso amplificando) los sesgos inherentes al *dataset*, limitando tanto la capacidad discriminativa de los modelos como la validez de sus evaluaciones.

En tercer lugar, este trabajo subraya la necesidad de clarificar la formulación de los problemas computacionales en arqueología. La categoría genérica “pintura rupestre”, heredada de un *dataset* heterogéneo, resulta excesivamente amplia para capturar la diversidad interna del fenómeno y restringe la capacidad de discriminación del modelo. La incorporación de categorías más específicas, anotaciones negativas, segmentaciones más detalladas o enfoques multimodales constituye una línea prioritaria para mejorar la pertinencia arqueológica de estos sistemas. Incluso en el caso de modelos abiertos y accesibles, la dependencia estructural de arquitecturas de *deep learning* impone límites que deben ser reconocidos explícitamente desde una perspectiva arqueológica,

especialmente en contextos caracterizados por una elevada complejidad visual y una fuerte variabilidad formal.

Desde una perspectiva arqueológica, los resultados muestran que la IA no reemplaza la interpretación experta, pero sí puede integrarse eficazmente en procesos de documentación, cribado inicial y prospección. Su principal utilidad reside en agilizar tareas repetitivas, explorar grandes volúmenes de imágenes y actuar como filtro preliminar que dirige la atención hacia zonas potencialmente relevantes. Su papel debe entenderse, por tanto, como el de un asistente analítico que complementa (no sustituye) la mirada experta, especialmente cuando se combina con documentación de alta calidad y análisis contextual.

Finalmente, creemos que lo expuesto refuerza la urgencia de avanzar hacia un marco metodológico común para la investigación en IA aplicada al arte rupestre. Ello implica desarrollar repositorios abiertos con anotaciones normalizadas y supervisión arqueológica directa, establecer protocolos de evaluación reproducibles y garantizar la liberación transparente de modelos y de sus parámetros entrenados. Solo mediante este enfoque integrado será posible valorar con precisión qué aportaciones pueden realizar estas tecnologías y cómo pueden incorporarse de forma responsable, rigurosa y eficaz a la investigación del arte rupestre.

## PÁGINAS WEB

Consejería de Educación, Cultura y Deportes de Castilla-La Mancha. <https://cultura.castillalamancha.es/>

Museo de la Valltorta. <https://museudelavalltorta.gva.es/es>

## BIBLIOGRAFÍA

- Bellat, M., Orellana, J. D., Tennie, C. y Scholten, T. (2025). Machine learning applications in archaeological practices: A review. *ArXiv*. 2501.03840v3. <https://doi.org/10.48550/arXiv.2501.03840>
- Bevan, A. (2015). The data deluge. *Antiquity* 89(348): 1473-1484. <https://doi.org/10.15184/aqy.2015.102>
- Bonald, L., Mützenberg, D., Krempser, E. y Verhagen, P. (2024). Predicting rock art sites in the Pajeú watershed, Brazil. *Digital Applications in Archaeology and Cultural Heritage* 35. e00372.
- Bruno, F., Bruno, S., De Sensi, G., Luchi, M. L., Mancuso, S. y Muzzupappa, M. (2013). From 3D reconstruction to virtual reality: A complete methodology for digital archaeological exhibition. *Journal of Cultural Heritage* 11(1): 42-49.

- Bruseker, G., Carboni, N. y Guillem, A. (2017). Cultural heritage data management: The role of formal ontology and CIDOC CRM. En Vincent, M. L., López-Mencheró Bendicho, V. M., Ioannides, M. y Levy, T. E. (eds.): *Heritage and archaeology in the digital age: Acquisition, curation, and dissemination of spatial cultural heritage data* (pp. 93-131). Springer, Cham.
- Casillo, M., F. Colace, R. Gaeta, A. Lorusso y M. Pellegrino (2024). Artificial intelligence in archaeological site conservation: Trends, challenges, and future directions. *Journal of Computer Applications in Archaeology* 8(1): 224-241. <https://doi.org/10.5334/jcaa.207>
- Chen, Y., Mancini, M., Zhu, X. y Akata, Z. (2024). Semi-supervised and unsupervised deep visual learning: A survey. *arXiv*. 2208.11296v1. <https://doi.org/10.48550/arXiv.2208.11296>
- Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Manning, Shelter Island.
- Cioni, D., Berlincioni, L., Becattini, F. y Del Bimbo, A. (2023). Diffusion based augmentation for captioning and retrieval in cultural heritage. *arXiv*. 2308.07151v1. <https://doi.org/10.48550/arXiv.2308.07151>
- De Lara, H. (2025). *Manifestaciones rupestres prehistóricas en el sur de Europa y el norte de África: el entorno del estrecho de Gibraltar* [Tesis doctoral]. UNED, Madrid. <https://e-spacio.uned.es/entidades/publication/1115c451-df2c-4621-b2ce-9d32f1ed56fc>
- De Lara, H., Mas, M. y Solís, M. (2025). Chronocultural proposal for the Atlanterra Cave (Cadiz, Spain). *Rock Art Research* 42(2): 213-233. <https://rockartresearch.com/bindex.php/rock/article/view/686>
- Dell'Unto, N., Landeschi, G., Apel, J. y Poggi, G. (2016). Experiencing ancient buildings from a 3D GIS perspective: A case drawn from the Swedish Pompeii project. *Journal of Archaeological Method and Theory* 23(1): 73-94.
- Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A. y James, S. (2020). Machine learning for cultural heritage: A survey. *Pattern Recognition Letters* 133: 102-108.
- Goodfellow, I., Bengio, Y. y Courville, A. (2016). *Deep learning*. MIT Press, Cambridge (MA).
- Guo, C., Pleiss, G., Sun, Y. y Weinberger, K. Q. (2017). On calibration of modern neural networks. *arXiv*. 1706.045992. <https://doi.org/10.48550/arXiv.1706.04599>
- Harth, A. (2024). The study of pigments in cultural heritage: A review using machine learning. *Heritage* 7(7): 3664-3695. <https://doi.org/10.3390/heritage7070174>
- He, K., Gkioxari, G., Dollár, P. y Girshick, R. (2017). Mask R-CNN. *arXiv*. 1703.06870v3. <https://doi.org/10.48550/arXiv.1703.06870>
- Historic England (2017). *Photogrammetric applications for cultural heritage*. Historic England Publishing, Swindon.
- Horn, C., Ivarsson, O., Lindhé, C., Potter, R., Ashely, G. y Ling, J. (2022). Artificial intelligence, 3D documentation, and rock art—Approaching and reflecting on the automation of identification and classification of rock art images. *Journal of Archaeological Method and Theory* 29: 188-213. <https://doi.org/10.1007/s10816-021-09518-6>
- Huo, D., Yang, S. y Hou, M. (2025). Using the improved YOLOv7-Seg model to segment symbols from rock art images. *npj Heritage Science* 13. 16. <https://doi.org/10.1038/s40494-025-01620-2>
- Jalandoni, A. (2021). An overview of remote sensing deliverables for rock art research. *Quaternary International* 572: 131-138. <https://doi.org/10.1016/j.quaint.2019.10.008>
- Jalandoni, A., Zhang, Y. y Zaidi, N. A. (2022). On the use of machine learning methods in rock art research with application to automatic painted rock art identification. *Journal of Archaeological Science* 144. 105629. <https://dx.doi.org/10.1016/j.jas.2022.105629>
- Kowlessar, J., Keal, J., Wesley, D., Moffat, I., Lawrence, D., Weson, A., Nayinggul, A. y Mimul Land Management Aboriginal Corporation (2021). Reconstructing rock art chronology with transfer learning: A case study from Arnhem Land, Australia. *Australian Archaeology* 87: 115-126. <https://doi.org/10.1080/03122417.2021.1895481>
- LeCun, Y., Bengio, Y. y Hinton, G. (2015). Deep learning. *Nature* 521(7553): 436-444. <https://doi.org/10.1038/nature14539>
- Liu, Y., Wang, Y. y Liu, C. (2023). A deep-learning method using auto-encoder and generative adversarial network for anomaly detection on ancient stone stele surfaces. *arXiv*. 2308.04426v1. <https://doi.org/10.48550/arXiv.2308.04426>
- Monna, F., Magail, J., Rolland, T., Navarro, N., Wilczek, J., Gantulga, J.-O., Esin, Y., Granjon, L., Allard, A.-C. y Chateau-Smith, C. (2020). Machine learning for rapid mapping of archaeological structures made of dry stones: Example of burial monuments from the Khirgisuur culture, Mongolia. *Journal of Cultural Heritage* 43: 118-128. <https://doi.org/10.1016/j.culher.2020.01.002>
- Pan, S. J. y Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10): 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- Ponmagal, R. S. y Srinivasan, N. (2015). Machine learning approach for exploring rock arts through the cloud infrastructure. En *Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (pp. 1-6). IEEE, Madurai (India). <https://doi.org/10.1109/ICCIC.2015.7435682>
- Redmon, J., Divvala, S., Girshick R. y Farhadi, A. (2016). You only look once: Unified, real-time object detection. *arXiv*. 1506.02640v5. <https://doi.org/10.1109/CVPR.2016.91>
- Roboflow Universe, 2024, *RMR Computer Vision Model*. <https://universe.roboflow.com/manifestaciones-y-gifuras-rupestres/rmr/model/1>
- Sabatelli, M., Kestemont, M., Daelemans, W., y Geurts, P. (2018). Deep transfer learning for art classification problems. En *Computer Vision - ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings, Part II* (pp. 631-646). Springer-Verlag, Berlin. [https://doi.org/10.1007/978-3-030-11012-3\\_48](https://doi.org/10.1007/978-3-030-11012-3_48)
- Schmarje, L., Santarossa, M., Schröder, S. M. y Koch, R. (2020). A survey on semi-, self- and unsupervised learning for image classification. *arXiv*. 2002.08721v5. <https://doi.org/10.48550/arXiv.2002.08721>
- Shorten, C. y Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data* 6(1). 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Sokolovska, A. (2024). Artificial intelligence and prospects for its use in the study of Upper Palaeolithic rock art. *Vita Antiqua* 15: 127-135. <https://doi.org/10.37098/VA-2024-15-127-135>
- Steels, L. y Wahle, B. (2020). Perceiving the focal point of a painting with AI: Case studies on works of Luc Tuymans. En Rocha, A., Steels, L. y van den Herik, J. (eds.): *Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART)*, Vol. 2 (pp. 895-901). ScitePress, Setúbal.



- Suhaimi, M. S., Zainuddin, K., Ghazali, M. D., Marzuki, F., Samad, A. M. y Majid, Z. (2023). Comparison of one-stage and two-stage strategies of machine learning model for rock art object detection. En *2023 IEEE 13th International Conference on System Engineering and Technology (ICSET)* (pp. 215-220). IEEE, Piscataway.
- Turner-Jones, R. N., Tuxworth, G., Haupt, R. A. y Wallis, L. (2024). Digitising the deep past: Machine learning for rock art motif classification in an educational citizen science application. *Journal on Computing and Cultural Heritage* 17(4): 1-19.
- Veggi, M. (2025). State of the art on artificial intelligence resources for interaction media design in digital cultural heritage. *arXiv*. 2504.13894v1. <https://doi.org/10.48550/arXiv.2504.13894>
- Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3. 160018. <https://doi.org/10.1038/sdata.2016.18>
- Yang, X., Song, Z., King, I. y Xu, Z. (2021). A survey on deep semi-supervised learning. *arXiv*. 2103.00550v2. <https://doi.org/10.48550/arXiv.2103.00550>
- Zeppelzauer, M., Poier, G., Seidl, M., Reinbacher, C., Schuster, S., Breiteneder, C. y Bischof, H. (2016). Interactive 3D segmentation of rock-art by enhanced depth maps and gradient preserving regularization. *Journal on Computing and Cultural Heritage* 9(4), Article 19: 1-30. <https://doi.org/10.1145/2950062>
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. y He, Q. (2019). A comprehensive survey on transfer learning. *arXiv*. 1911.02685v3. <https://doi.org/10.48550/arXiv.1911.02685>