



**DETECCIÓN AUTOMÁTICA DE BIOINDICADORES DE DETERIORO EN
IMÁGENES DE ARTE RUPESTRE DE LA SERRANÍA DE LA LINDOSA
MEDIANTE TECNICAS DE APRENDIZAJE AUTOMÁTICO.**

Luis Alberto Echeverry Henao y Jhon Sebastian Rengifo Segura

Director:
PhD. Manuel Guillermo Forero Vargas

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, JUNIO 22 DE 2026

Agradecimientos.

Dedicamos este logro, en primer lugar, a nuestros padres, por ser pilares de amor, resiliencia y responsabilidad en nuestras vidas. Agradecemos también a Claudia Echeverry por su apoyo incondicional y sabias orientaciones. Nuestro sincero reconocimiento al Dr. Manuel Forero por la paciencia, guía y disposición brindadas en este proceso. Asimismo, expresamos nuestra gratitud a Cesar Yair Ardila Peña, arqueólogo del ICANH, cuya gestión hizo posible el desarrollo de este proyecto de investigación.

TABLA DE CONTENIDO

1.	INTRODUCCIÓN.....	6
2.	DEFINICIÓN DEL PROBLEMA.....	8
2.1	PLANTEAMIENTO DEL PROBLEMA.....	8
2.2	FORMULACIÓN DEL PROBLEMA	10
3.	OBJETIVOS DEL PROYECTO.....	11
3.1	OBJETIVO GENERAL.....	11
3.2	OBJETIVOS ESPECÍFICOS	11
4.	MARCO DE REFERENCIA	12
4.1	MARCO TEÓRICO	12
4.1.1	Deterioro biológico en arte rupestre y bioindicadores	12
4.1.2	Visión por computador y procesamiento de imágenes patrimoniales	12
4.1.3	Aprendizaje automático y segmentación de imágenes.....	13
4.1.4	Aprendizaje automático aplicado al diagnóstico visual.....	14
4.1.5	Evaluación de modelos y métricas de desempeño	14
4.1.6	Estado del arte y tendencias interdisciplinarias	14
4.1.7	Arquitectura U-Net.....	15
4.1.8	Data leakage y validación	15
4.2	ANTECEDENTES.....	16
5.	METODOLOGÍA	18
5.1	Selección de los deterioros objeto de estudio.....	18
5.2	Construcción de máscaras binarias.....	21
5.3	Construcción del conjunto de datos para entrenamiento.....	26
5.3.1	Generación de <i>patches</i>	26
5.3.2	Parámetros para la generación de patches	27
5.3.3	Organización y trazabilidad del conjunto de datos	29
5.3.4	Estrategia de partición y validación experimental	29
6.	DISEÑO E IMPLEMENTACIÓN DEL MODELO	33
6.1	Flujo de trabajo para entrenamiento del modelo	33
6.2	Arquitecturas evaluadas	34
6.3	Configuración experimental.....	34
7.	RESULTADOS	36
7.1	Resultados obtenidos para el deterioro plantas superiores.....	36

7.1.1	Evaluación del modelo U-Net sin balanceo de clases	36
7.1.2	Evaluación del modelo U-Net con estrategia de undersampling adaptativo.....	44
8.	DISCUSIÓN.....	53
8.1	Problemas computacionales y limitaciones técnicas.....	54
9.	CONCLUSIONES Y TRABAJOS FUTUROS	56
9.1	Conclusiones	56
9.2	Trabajos Futuros	58
10.	Anexos.....	61

LISTA DE TABLAS

Tabla 1.	Metadatos suministrados por el ICANH	20
Tabla 2.	Descripción de los principales campos utilizados en los archivos de metadatos de los patches generados para entrenamiento y validación.	29
Tabla 3.	Hiperparámetros iniciales seleccionados para el entrenamiento.	35

LISTA DE FIGURAS

Figura 1.	Representación 2D Panel Dantas.	18
Figura 2.	Mapa de deterioro Panel Dantas.	19
Figura 3.	Flujo metodológico para la generación de máscaras binarias a partir de los mapas de deterioro suministrados por el ICANH.	21
Figura 4.	Ejemplo de Factor biológico plantas superiores en mapa de deterioro.....	22
Figura 5.	Delimitación del contorno usando polygon selection de ImageJ.....	23
Figura 6.	Mascara binaria resultante.	24
Figura 7.	Validación visual de regiones segmentadas correspondientes a deterioros estructurales tipo galerías en el panel PC_CRLTO_NT.....	25
Figura 8.	Proceso de generación de máscara binaria para deterioros tipo galerías en el panel PC_PP_NT mediante segmentación manual en ImageJ.	25
Figura 9.	Ejemplo de segmentación manual del deterioro tipo galerías en el panel PC_DNT_CA, incluyendo imagen original, máscara binaria y visualización overlay.	25
Figura 10.	Ejemplo de patch positivo generado para entrenamiento del modelo de segmentación semántica, incluyendo imagen original (a), máscara binaria (b) y visualización overlay del deterioro identificado (c).	27
Figura 11.	Detalle de patch resultante con dimensión 128×128 para galerías.	28
Figura 12.	Proceso de validación usando LOIO.....	31
Figura 13.	Distribución de métricas Dice, IoU y HD95 obtenidas durante la evaluación del modelo U-Net.	37

Figura 14. Ejemplos de mejores predicciones del modelo U-Net según la métrica HD95 para la clase plantas superiores.	39
Figura 15. Ejemplos de mejores predicciones del modelo U-Net según la métrica Dice para la clase plantas superiores.	40
Figura 16. Ejemplos de predicciones por parche del modelo U-Net, incluyendo imagen original, máscara real, máscara predicha y superposición.	41
Figura 17. Ejemplos de peores predicciones del modelo U-Net según la métrica HD95 para la clase plantas superiores.	42
Figura 18. Ejemplos de peores predicciones del modelo U-Net según la métrica Dice para la clase plantas superiores.	43
Figura 19. Histogramas y diagramas de caja de las métricas Dice, IoU y HD95 obtenidas durante la evaluación del modelo U-Net con undersampling adaptativo.	45
Figura 20. Ejemplos de predicciones por parche del modelo U-Net con undersampling adaptativo, incluyendo imagen original, máscara real, máscara predicha y superposición.	46
Figura 21. Ejemplos de mejores predicciones del modelo U-Net con undersampling adaptativo según la métrica HD95 para la clase plantas superiores.	48
Figura 22. Ejemplos de mejores predicciones del modelo U-Net con undersampling adaptativo según la métrica Dice para la clase plantas superiores.	49
Figura 23. Ejemplos de peores predicciones del modelo U-Net con undersampling adaptativo según la métrica HD95 para la clase plantas superiores.	50
Figura 24. Ejemplos de peores predicciones del modelo U-Net con undersampling adaptativo según la métrica Dice para la clase plantas superiores.	51

INTRODUCCIÓN

El patrimonio arqueológico y cultural constituye uno de los elementos más importantes para la preservación de la memoria histórica de las sociedades. En Colombia, la Serranía de La Lindosa, ubicada en el departamento del Guaviare, representa uno de los complejos de arte rupestre más relevantes del país debido a la gran cantidad de representaciones pictográficas presentes en sus paneles rocosos. Además, este conjunto de pinturas rupestres constituyen uno de los sitios arqueológicos más importantes de América Latina, con registros pictográficos que superan los 12.500 años de antigüedad según estudios recientes [1, pp. 14-39], [2]. Este patrimonio cultural posee un valor excepcional debido a su relevancia histórica, simbólica y antropológica, permitiendo comprender dinámicas sociales y territoriales de comunidades prehispánicas que habitaron la Amazonía colombiana.

Sin embargo, estos registros arqueológicos se encuentran expuestos a múltiples factores de deterioro biológico y ambiental que afectan progresivamente su conservación y dificultan los procesos de monitoreo realizados por expertos. En los resultados de las diferentes investigaciones y reportes técnicos, se ha evidenciado que gran parte de estos paneles presentan procesos de deterioro asociados a factores biológicos, ambientales y antrópicos. Entre los principales agentes identificados se encuentran biofilms, hongos, líquenes, carbonataciones, microfisuras y crecimiento de plantas superiores, los cuales afectan progresivamente la estabilidad física y visual de las superficies pictográficas [3, pp. 67-76], [4].

Actualmente, los procesos de diagnóstico y monitoreo dependen principalmente de inspecciones visuales de formar presencial realizadas por especialistas, que permiten documentar manualmente las condiciones de deterioro de los paneles, generando procesos extensos, subjetivos y difíciles de escalar en contextos geográficos complejos [5]. Adicionalmente, las condiciones ambientales y los conflictos territoriales que hacen parte del contexto general del departamento del Guaviare, complejizan el acceso a muchos de los paneles, lo que limita la frecuencia de documentación y seguimiento y dificulta la implementación de estrategias preventivas de conservación.

Ante esta problemática, el presente proyecto propone el desarrollo de un modelo basado en aprendizaje profundo y visión por computador orientado a la detección automática de bioindicadores de deterioro presentes en imágenes digitales de alta resolución. La metodología planteada integra técnicas de procesamiento digital de imágenes, construcción manual de máscaras binarias, generación de patches y entrenamiento de arquitecturas de segmentación semántica como U-Net y ResU-Ne, ampliamente utilizadas en problemas de clasificación pixel a pixel [6]. El proyecto fue desarrollado utilizando imágenes suministradas por el Instituto Colombiano de Antropología e Historia (ICANH) [7], las cuales requirieron procesos complejos de preprocesamiento debido a su alta resolución y a la necesidad de construir anotaciones supervisadas manualmente.

Asimismo, durante el desarrollo experimental se identificaron múltiples retos relacionados con la capacidad computacional, el almacenamiento y los tiempos de entrenamiento, lo que motivó la búsqueda de alternativas de aceleración mediante *GPU* y plataformas externas. A pesar de estas dificultades, el proyecto permitió consolidar un flujo de trabajo funcional de segmentación semántica y establecer bases metodológicas sólidas para futuras aplicaciones de *Machine Learning* en conservación patrimonial, alineándose además con tendencias internacionales relacionadas con IA aplicada al patrimonio cultural y ciencia abierta promovidas por la UNESCO [8].

Los resultados obtenidos mediante el entrenamiento de los modelos con arquitectura U-Net y ResU-Net para las clases plantas superiores, galerías, nidos y nidos de avispa, permitieron identificar avances relevantes y, al mismo tiempo, retos metodológicos importantes. En particular, se evidenció que el desbalance entre *patches* con y sin deterioro influye de manera significativa en la interpretación de las métricas globales, por lo que fue necesario evaluar estrategias de balanceo como el *undersampling* adaptativo. Además, se pudo identificar que el tamaño de los *patches* influye de manera significativa en los resultados obtenidos, siendo una variable determinante al momento de realizar la construcción de *dataset* con el que se entrenó y probó cada uno de los modelos. Estos hallazgos fortalecen la comprensión del problema y permiten establecer una ruta experimental más clara para futuros proyectos, orientados a mejorar la detección real de deterioros y la generalización de este modelo en escenarios patrimoniales.

DEFINICIÓN DEL PROBLEMA

2.1 PLANTEAMIENTO DEL PROBLEMA

Las pinturas rupestres de la Serranía de La Lindosa representan un legado arqueológico de alto valor histórico y cultural para Colombia y América Latina [1], [2]. Estas manifestaciones gráficas contienen información relevante sobre prácticas simbólicas y dinámicas territoriales de comunidades prehispánicas, las cuales constituyen una fuente invaluable para las investigaciones arqueológicas contemporánea, que, a su vez permiten el estudio de las sociedades prehispánicas y la preservación del patrimonio nacional. Sin embargo, los paneles rupestres enfrentan procesos continuos de deterioro asociados tanto a factores naturales como antrópicos. Entre los agentes más relevantes se encuentran la humedad, la erosión, la radiación solar, la proliferación de microorganismos y el crecimiento de biofilms, líquenes y vegetación superior [2], [4]. Asimismo, actividades humanas como vandalismo, el turismo no regulado y el contacto físico directo, incrementan el deterioro de las superficies pictográficas [9].

Aunque el ICANH ha desarrollado procesos de documentación visual y generación de mapas de deterioro, dichos procedimientos continúan dependiendo de análisis manuales realizados por expertos, implicando altos costos temporales y operativos. La clasificación de deterioros en un único panel puede requerir semanas de trabajo especializado, dificultando la implementación de un monitoreo continuo a gran escala. Desde el punto de vista tecnológico, existe un vacío importante en la aplicación de herramientas de aprendizaje profundo orientadas específicamente a la segmentación automática de deterioros en arte rupestre colombiano. Si bien modelos de segmentación semántica como U-Net y DeepLabv3+ han mostrado resultados sobresalientes en áreas como imágenes biomédicas y análisis satelital [6], su aplicación en patrimonio cultural aún es limitada, especialmente en contextos latinoamericanos.

Por tanto, el problema central del proyecto radica en la ausencia de herramientas con bases sólidas en ciencia de datos, que involucren *Machine Learning* o *Deep Learning* y que permitan detectar y segmentar bioindicadores de deterioro en imágenes de arte rupestre de manera eficiente, reproducible y escalable, contribuyendo así a procesos de monitoreo y conservación patrimonial no invasiva. Actualmente, el monitoreo y análisis de estos deterioros se realiza principalmente de manera manual por parte de expertos, mediante inspección visual, pero presentando limitaciones importantes relacionadas con el tiempo requerido, la subjetividad en la interpretación y la dificultad de procesar grandes volúmenes de información visual.

La aplicación de técnicas de aprendizaje profundo para el monitoreo del deterioro en bienes patrimoniales enfrenta desafíos asociados al procesamiento de imágenes de alta resolución y a la disponibilidad de conjuntos de datos adecuadamente etiquetados. En el contexto del arte rupestre, la construcción de máscaras de segmentación requiere un alto nivel de intervención especializada, lo que representa un gran reto en especial si no se cuenta con un grupo de expertos. Adicionalmente, el manejo de imágenes de alta resolución representa un desafío técnico significativo. Las imágenes disponibles de los paneles rupestres tienen dimensiones aproximadas de 15360×8640 píxeles, lo que implica altos requerimientos de memoria, almacenamiento y capacidad computacional para su procesamiento.

Frente a este panorama, surge la necesidad de desarrollar herramientas basadas en técnicas de inteligencia artificial y visión por computador que permitan apoyar la identificación, detección y segmentación de bioindicadores de deterioro presentes en imágenes de arte rupestre. La automatización parcial de estos procesos puede contribuir a mejorar la eficiencia del monitoreo, reducir la dependencia de procedimientos manuales y proporcionar información objetiva que apoye la toma de decisiones orientadas a la conservación del patrimonio arqueológico.

2.2 FORMULACIÓN DEL PROBLEMA

¿Cómo desarrollar un modelo basado en aprendizaje profundo capaz de detectar y clasificar automáticamente los principales bioindicadores de deterioro en imágenes de paneles de arte rupestre de la Serranía de La Lindosa, con el fin de apoyar su monitoreo y conservación?

- ¿Qué técnicas de segmentación, normalización y aumento de datos son más adecuadas para preprocesar las imágenes suministradas por el ICANH y construir un conjunto de datos balanceado para entrenamiento?
- ¿Cuáles son las características visuales y morfológicas más relevantes de los bioindicadores de deterioro presentes en las imágenes segmentadas, y cómo pueden ser utilizadas como insumos para el entrenamiento del modelo?
- ¿Qué tipo de arquitectura de aprendizaje profundo es más adecuada para segmentar y clasificar automáticamente las regiones afectadas por deterioro en las imágenes de paneles rupestres?
- ¿Qué nivel de desempeño presenta el modelo propuesto en términos de métricas como precisión, sensibilidad, especificidad, F1-score e IoU al clasificar las regiones deterioradas?
- ¿Qué tan efectiva resulta la herramienta desarrollada al aplicarse en estudios de caso reales, como sistema de apoyo no invasivo para el monitoreo y conservación del arte rupestre?

OBJETIVOS DEL PROYECTO

3.1 OBJETIVO GENERAL

Desarrollar un modelo basado en aprendizaje profundo para detectar y clasificar automáticamente los principales bioindicadores de deterioro en imágenes de paneles de arte rupestre de la Serranía de La Lindosa, con el propósito de apoyar el monitoreo y la conservación del patrimonio arqueológico.

3.2 OBJETIVOS ESPECÍFICOS

- Preprocesar las imágenes suministradas por el ICANH mediante técnicas de segmentación, normalización y aumento de datos, con el fin de construir un conjunto de datos balanceado y adecuado para el entrenamiento del modelo.
- Caracterizar visual y morfológicamente los bioindicadores de deterioro presentes en las regiones segmentadas, utilizando la información previamente anotada en los mapas de deterioro suministrados por el ICANH.
- Diseñar un modelo de aprendizaje profundo orientado a la segmentación semántica y clasificación de regiones afectadas por deterioro, adaptado a las características del conjunto de datos.
- Validar el modelo utilizando métricas de desempeño como precisión, sensibilidad, especificidad, F1-score e IoU, sobre un conjunto de datos anotado por expertos.
- Analizar la utilidad del modelo como herramienta de apoyo para el monitoreo del estado de conservación, mediante pruebas de desempeño en estudios de caso seleccionados.

MARCO DE REFERENCIA

4.1 MARCO TEÓRICO

El desarrollo de soluciones basadas en inteligencia artificial para el análisis y conservación del patrimonio cultural requiere la integración de conocimientos provenientes de diferentes disciplinas. En el contexto del presente estudio, resulta necesario comprender tanto los procesos de deterioro que afectan el arte rupestre como los fundamentos de las técnicas de visión por computador y aprendizaje profundo empleadas para su análisis automatizado. Este marco teórico reúne los conceptos fundamentales relacionados con la conservación del patrimonio arqueológico, los bioindicadores de deterioro presentes en los paneles rupestres, el procesamiento digital de imágenes y las técnicas de segmentación semántica basadas en redes neuronales profundas. Asimismo, se presentan las principales arquitecturas y métricas de evaluación utilizadas en problemas de segmentación de imágenes, las cuales constituyen la base conceptual para el desarrollo de la investigación.

4.1.1 Deterioro biológico en arte rupestre y bioindicadores

El arte rupestre es altamente vulnerable a procesos de deterioro, en particular al biodeterioro, causado por el crecimiento de microorganismos como líquenes, musgos, algas y bacterias, que se agrupan en biofilms. Estos organismos generan transformaciones químicas y físicas que degradan los pigmentos y la roca soporte. Según Caneva et al., el biodeterioro en patrimonio cultural puede entenderse como la “alteración de materiales por organismos vivos, directa o indirectamente”[10]. Las condiciones del clima húmedo del Guaviare, junto con la exposición solar intermitente, crean un microambiente propicio para la proliferación de estos agentes. Los bioindicadores son utilizados no sólo como síntomas del deterioro, sino también como indicadores ambientales que ayudan a entender las condiciones que aceleran el daño. Documentarlos de forma sistemática es clave para establecer estrategias de conservación preventiva.

4.1.2 Visión por computador y procesamiento de imágenes patrimoniales

El procesamiento digital de imágenes consiste en transformar y analizar imágenes mediante algoritmos computacionales. En el contexto de conservación patrimonial, estas técnicas permiten resaltar patrones, texturas y estructuras que pueden ser indicadores de daño o cambio en el tiempo. Entre las transformaciones más empleadas están el aumento de datos (data augmentation), la normalización de color, y los filtros adaptativos para mejorar la visibilidad de características degradadas. Estudios como el de Bayarri et al. han utilizado técnicas de detección multispectral para analizar la superficie de pinturas rupestres, combinando diferentes longitudes de onda para aislar contaminantes o alteraciones invisibles al ojo humano [11].

4.1.3 Aprendizaje automático y segmentación de imágenes

El aprendizaje automático (machine learning, ML) comprende un conjunto de técnicas algorítmicas que permiten a los sistemas computacionales aprender patrones a partir de los datos y realizar predicciones o tomar decisiones sin haber sido programados explícitamente para cada escenario. Dentro del análisis de imágenes, los enfoques clásicos de aprendizaje supervisado, como las Máquinas de Vectores de Soporte (Support Vector Machines, SVM) y los algoritmos basados en árboles de decisión como Random Forest, han demostrado resultados favorables en problemas donde el volumen de datos es limitado y las características visuales pueden ser extraídas y representadas manualmente mediante descriptores previamente definidos [12], [13].

Sin embargo, cuando el objetivo no consiste únicamente en clasificar una imagen completa sino en identificar y delimitar regiones específicas dentro de ella por ejemplo, zonas afectadas por bioindicadores de deterioro sobre superficies rupestres resulta más apropiado emplear técnicas de segmentación semántica. Este tipo de modelos busca asignar una etiqueta a cada píxel de la imagen, permitiendo construir una representación espacial detallada del objeto o fenómeno de interés. A diferencia de los métodos tradicionales de clasificación, la segmentación requiere preservar relaciones espaciales y estructuras locales presentes en la imagen, lo que incrementa significativamente la complejidad del problema [14].

Dentro de los enfoques más utilizados para resolver problemas de segmentación semántica se encuentran las Redes Neuronales Convolucionales (*Convolutional Neural Networks*, CNN), las cuales han mostrado una alta capacidad para aprender representaciones jerárquicas directamente desde los datos de entrada. Estas arquitecturas utilizan operaciones de convolución para extraer características de diferentes niveles de abstracción, permitiendo reconocer patrones complejos relacionados con forma, textura y contexto espacial sin necesidad de definir manualmente los atributos visuales relevantes [13], [14].

Modelos basados en CNN como U-Net, DeepLabV3+ y Mask R-CNN han sido ampliamente utilizados en aplicaciones de segmentación en áreas como biomedicina, agricultura, inspección automática y diagnóstico no invasivo. U-Net introdujo una arquitectura tipo *encoder-decoder* con conexiones de salto (*skip connections*) que permiten conservar información espacial durante el proceso de reconstrucción de la máscara segmentada. Por su parte, DeepLabV3+ incorporó convoluciones dilatadas y mecanismos avanzados de recuperación espacial para mejorar la precisión de segmentación en objetos complejos [15], mientras que Mask R-CNN extendió los modelos de detección de objetos incorporando segmentación a nivel de instancia [16].

Aunque el uso de estas técnicas en contextos patrimoniales aún es relativamente reciente, su adopción ha venido aumentando debido a su potencial para automatizar procesos tradicionalmente manuales, como la identificación, documentación y monitoreo del deterioro sobre bienes culturales. En particular, la aplicación de modelos de segmentación sobre imágenes patrimoniales permite generar herramientas de apoyo para procesos de conservación preventiva y análisis no invasivo del patrimonio cultural [17].

4.1.4 Aprendizaje automático aplicado al diagnóstico visual

El machine learning supervisado permite entrenar algoritmos que, a partir de ejemplos etiquetados, aprenden a clasificar o segmentar nuevas imágenes. En problemas de diagnóstico visual, los modelos clásicos como *Support Vector Machines* (SVM) y Random Forest han mostrado buenos resultados cuando se cuenta con características estructuradas y conjuntos de datos moderados. No obstante, tareas como la identificación pixel a pixel de zonas de deterioro requieren enfoques más complejos como los basados en redes neuronales convolucionales (CNN), las cuales aprenden representaciones jerárquicas directamente de las imágenes. Modelos como U-Net, DeepLabV3+ y Mask R-CNN han sido ampliamente usados en aplicaciones médicas, detección de cultivos, mapeo satelital y conservación digital, por su capacidad para segmentar con precisión regiones complejas y amorfas.

4.1.5 Evaluación de modelos y métricas de desempeño

La evaluación de modelos de clasificación o segmentación debe considerar métricas robustas como:

Precisión (accuracy): tasa de aciertos global.

Recall (sensibilidad): qué tanto se logra detectar lo que realmente está deteriorado.

F1-score: equilibrio entre precisión y recall.

IoU (Intersection over Union): superposición entre la predicción y la anotación real.

Estas métricas permiten determinar si un modelo es aplicable a contextos reales y si tiene utilidad práctica en conservación.

4.1.6 Estado del arte y tendencias interdisciplinarias

En los últimos cinco años, se ha fortalecido el uso de inteligencia artificial en la protección del patrimonio cultural. Proyectos como TIME Machine EU, AI4Heritage, o los análisis de monumentos en Petra, han demostrado que los modelos de deep learning pueden integrarse a flujos de trabajo arqueológicos y museológicos [15].

En Latinoamérica, iniciativas como la de la Universidad de São Paulo y la Universidad de Buenos Aires han comenzado a aplicar modelos computacionales para monitorear cambios en templos coloniales y arte rupestre, aunque con enfoques todavía exploratorios [16].

Este proyecto contribuye a esta tendencia desde un enfoque interdisciplinar, alineado con las recomendaciones de UNESCO sobre ciencia abierta, uso ético de la IA y acceso universal al conocimiento para la conservación del patrimonio.

4.1.7 Arquitectura U-Net

Dentro de las arquitecturas de segmentación más utilizadas se encuentra U-Net, propuesta inicialmente para imágenes biomédicas y posteriormente aplicada en múltiples dominios. Esta arquitectura se caracteriza por una estructura encoder-decoder que permite capturar tanto información contextual como detalles espaciales mediante conexiones de salto (skip connections).

La arquitectura U-Net fue seleccionada como modelo base del proyecto debido a su eficiencia en tareas de segmentación binaria y multicategoría, especialmente en datasets limitados y problemas donde las regiones de interés presentan tamaños variables y formas irregulares.

4.1.8 Data leakage y validación

Uno de los principales retos metodológicos del proyecto estuvo relacionado con el riesgo de *data leakage*, debido a la limitada cantidad de paneles disponibles y a la alta similitud espacial entre *patches* provenientes de una misma imagen. El *data leakage* ocurre cuando durante el entrenamiento se incorpora, directa o indirectamente, información que posteriormente estará presente en el conjunto de validación o prueba, generando estimaciones artificialmente optimistas del desempeño del modelo y afectando su capacidad de generalización sobre datos no observados [18].

Este problema adquiere especial relevancia en tareas de segmentación semántica basadas en imágenes de alta resolución, donde la división de una misma imagen en múltiples *patches* puede generar dependencia espacial entre muestras aparentemente independientes. Cuando regiones cercanas del mismo panel son distribuidas simultáneamente entre entrenamiento y validación, el modelo puede aprender patrones específicos del contexto local y no necesariamente características generalizables del fenómeno estudiado, produciendo una sobreestimación de métricas de desempeño [19].

Como estrategia inicial se propuso implementar una metodología Leave-One-Image-Out Cross Validation (LOIO), donde un panel completo se utiliza para validación mientras los restantes son empleados para entrenamiento. Este enfoque busca reducir la dependencia espacial entre conjuntos y aproximar un escenario más realista de generalización. Sin embargo, dado que algunos tipos de deterioro no están presentes en todos los paneles disponibles, esta metodología presentó limitaciones importantes relacionadas con el desbalance entre clases y la pérdida de representatividad experimental [20].

Por esta razón, se evaluaron configuraciones de validación intra-panel controlada, procurando minimizar la contaminación entre conjuntos y mantener una separación metodológicamente consistente entre entrenamiento, validación y prueba. Esta estrategia buscó equilibrar la disponibilidad limitada de datos con la necesidad de obtener una evaluación experimental más robusta y representativa del desempeño real del modelo.

4.2 ANTECEDENTES

Entre los principales antecedentes de esta investigación se encuentran los procesos de documentación, conservación y caracterización desarrollados por el Instituto Colombiano de Antropología e Historia (ICANH) en las zonas arqueológicas protegidas de la Serranía de La Lindosa. Particularmente relevante resulta la colaboración realizada entre el ICANH, la Escuela Nacional de Conservación, Restauración y Museografía "Manuel del Castillo Negrete" y el Instituto Nacional de Antropología e Historia (INAH) de México, de la cual se derivó el conjunto de imágenes y mapas de deterioro empleados como insumo principal en el presente estudio [7].

El proceso de obtención de estos productos digitales consistió en la captura fotográfica sistemática de los paneles rupestres, complementada mediante la implementación de un sistema de referencia local basado en una estación total. Este instrumento permitió determinar con precisión las coordenadas espaciales de diferentes puntos de referencia sobre los paneles, facilitando la construcción de una cuadrícula de 5 m × 5 m utilizada para organizar y segmentar el área de estudio. Posteriormente, equipos interdisciplinarios de arqueólogos y conservadores identificaron y registraron manualmente los distintos tipos de deterioro sobre láminas de acetato transparente superpuestas a las impresiones de los paneles. Finalmente, estas anotaciones fueron digitalizadas y transformadas en mapas de deterioro georreferenciados, constituyendo un conjunto de datos estructurado que actualmente representa una de las principales fuentes de información para estudios relacionados con la conservación del arte rupestre.

Un antecedente particularmente cercano a la presente investigación corresponde al trabajo titulado *Análisis de la integración SIG, topografía y digitalización vectorial de imágenes para la gestión sostenible de la conservación y documentación de arte rupestre en el área arqueológica protegida Serranía de la Lindosa*. En esta investigación desarrollada en 2024, Rodríguez propuso una metodología basada en la integración de Tecnologías de la Información Geográfica (SIG), técnicas topográficas y procesos de digitalización vectorial para apoyar la documentación y conservación del patrimonio arqueológico. El estudio empleó tecnologías no invasivas como escáneres láser, receptores GPS y herramientas especializadas de procesamiento digital para generar modelos tridimensionales y productos gráficos destinados al análisis y monitoreo de los paneles rupestres.

La investigación se desarrolló sobre los paneles Corralito y Raudal de la Serranía de La Lindosa, demostrando el potencial de las tecnologías geoespaciales para fortalecer los procesos de documentación, conservación y gestión del patrimonio cultural. Asimismo, los productos digitales obtenidos constituyen un antecedente importante para investigaciones posteriores orientadas al análisis automatizado de información arqueológica. En este contexto, la presente investigación complementa dichos esfuerzos al incorporar técnicas de visión por computador y aprendizaje profundo para la detección y segmentación automática de deterioros biológicos presentes en los paneles rupestres [21].

Desde la perspectiva de la conservación patrimonial, también resulta relevante el trabajo desarrollado por Urzì y De Leo en 2020, quienes realizaron una revisión sobre los procesos de biodeterioro que afectan bienes culturales y arqueológicos. Los autores documentan cómo diferentes microorganismos, biofilms y formas de colonización biológica generan alteraciones físicas y químicas sobre los materiales patrimoniales, proporcionando criterios científicos para la identificación y caracterización de patrones visuales asociados al deterioro. Estos aportes constituyen una referencia importante para comprender los procesos de degradación biológica presentes en los paneles rupestres y respaldan la selección de los tipos

de deterioro abordados en la presente investigación [22].

A nivel nacional, la investigación desarrollada por Castiblanco en 2020 sobre el arte rupestre de la Serranía de La Lindosa aporta elementos fundamentales para comprender la relevancia cultural y arqueológica de este patrimonio, así como las amenazas que enfrenta actualmente. Dicho trabajo documenta diferentes factores de deterioro, tanto naturales como antrópicos, y resalta la necesidad de implementar estrategias de conservación preventiva que permitan monitorear de manera continua el estado de los paneles sin generar intervenciones invasivas sobre las superficies arqueológicas [1].

En conjunto, los antecedentes revisados evidencian importantes avances en la documentación digital del patrimonio rupestre, la caracterización de procesos de biodeterioro y la aplicación de tecnologías geoespaciales para la conservación arqueológica. Sin embargo, se identificó una limitada aplicación de técnicas de visión por computador y aprendizaje profundo orientadas a la detección automática de deterioros sobre los insumos documentales generados para la Serranía de La Lindosa. Esta situación constituye una oportunidad de investigación, dado que los productos derivados de los procesos de documentación realizados por el ICANH proporcionan una base adecuada para explorar estrategias de segmentación semántica que contribuyan al monitoreo y conservación no invasiva del patrimonio arqueológico.

METODOLOGÍA

La metodología propuesta fue diseñada para dar cumplimiento a los objetivos específicos de la investigación mediante una secuencia de actividades que abarcan la selección de los deterioros a segmentar, la construcción del conjunto de máscaras binarias, la preparación de los datos, el desarrollo de modelos de segmentación semántica y la evaluación de su desempeño. Cada una de las etapas metodológicas se encuentra alineada con los objetivos planteados, permitiendo establecer una relación directa entre las actividades ejecutadas y los resultados obtenidos durante el desarrollo del estudio.

5.1 Selección de los deterioros objeto de estudio.

Dentro de los insumos suministrados por el ICANH, se contó con dos tipos de imágenes fundamentales para el desarrollo del proyecto. El primer grupo corresponde a las representaciones bidimensionales (2D) de los paneles rupestres, cada uno estructurado con una cuadrícula de referencia de 5 m x 5 m (Figura 1). El segundo grupo está conformado por los mapas de deterioro, los cuales incluyen la simbología y leyendas correspondientes a cada uno de los tipos de alteración identificados por los arqueólogos durante el proceso de documentación en campo (Figura 2).

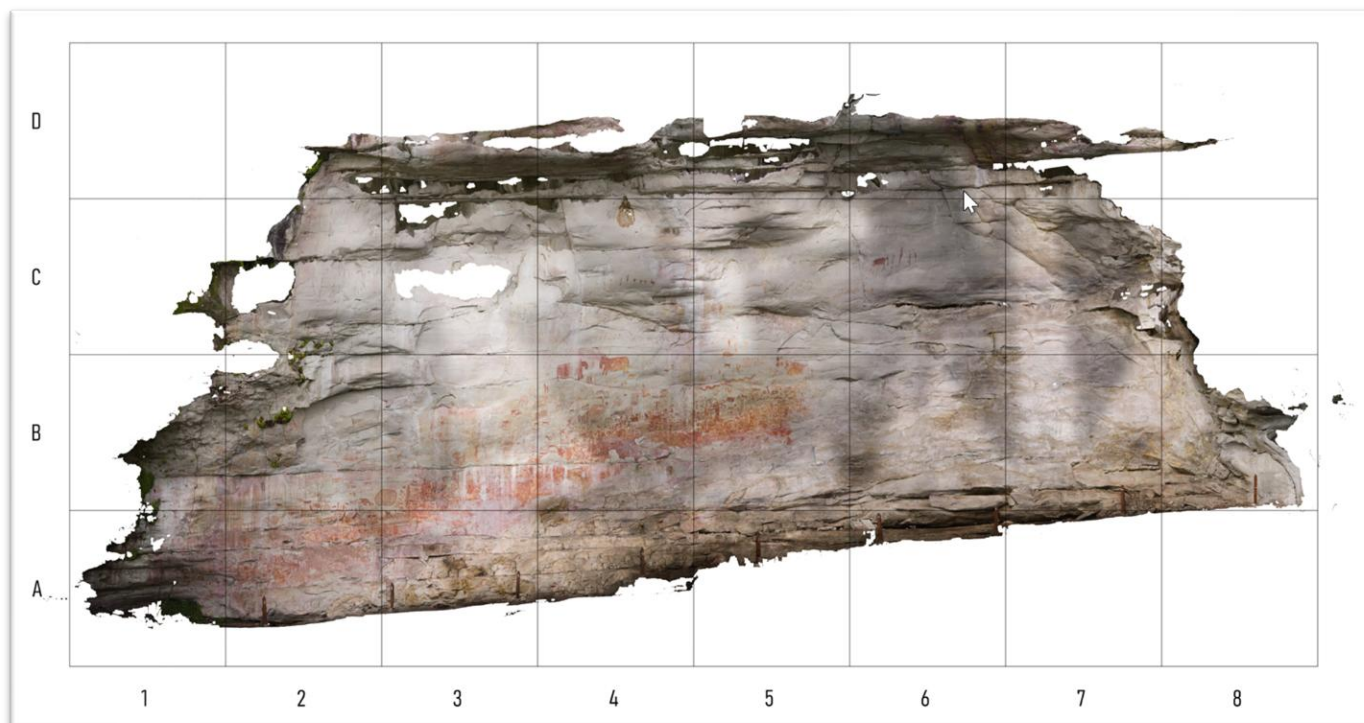


Figura 1. Representación 2D Panel Dantas.

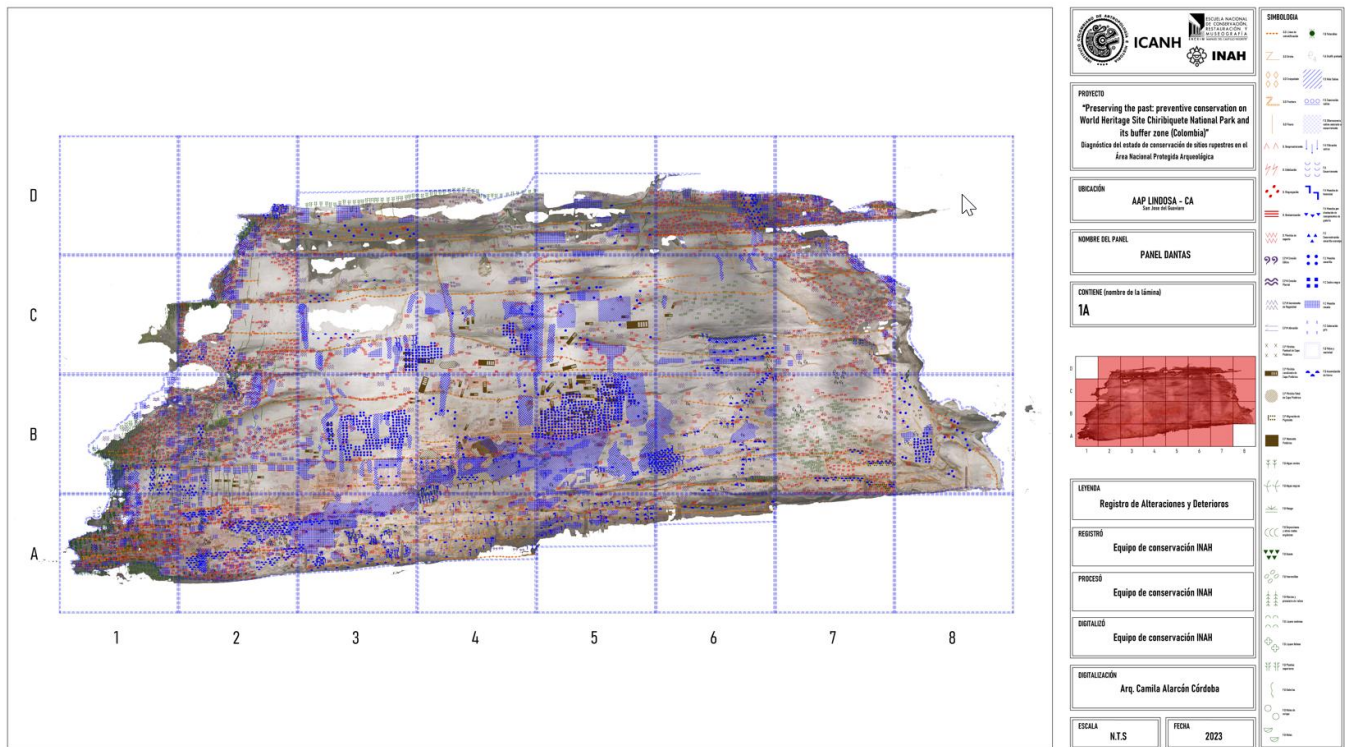


Figura 2. Mapa de deterioro Panel Dantas.

Bajo este esquema de documentación, el ICANH suministró información correspondiente a seis paneles de la Serranía de La Lindosa. Cuatro de ellos —Panel Dantas, Panel Corralito Nuevo Tolima, Panel Principal Cerro Azul y Panel Principal Nuevo Tolima— contaban tanto con las representaciones bidimensionales (2D) como con sus respectivos mapas de deterioro. Por su parte, para los paneles Raudal Guayabero y Largo Cerro Azul únicamente se dispuso de las representaciones 2D, sin los mapas de deterioro asociados. Debido a la ausencia de estas anotaciones especializadas, no fue posible generar las máscaras binarias requeridas para el entrenamiento y evaluación de los modelos de segmentación semántica. En consecuencia, dichos paneles fueron excluidos del desarrollo experimental, dado que no contaban con la información necesaria para establecer las regiones de referencia utilizadas durante el proceso de aprendizaje supervisado. Adicionalmente, el ICANH suministró un conjunto de metadatos asociados a cada imagen, incluyendo información como el nombre del panel documentado, su abreviatura, el tipo de procesamiento aplicado, la extensión del archivo, la resolución espacial y la fecha de generación. Un ejemplo de estos metadatos se presenta en la Tabla 1.

Tabla 1. Metadatos suministrados por el ICANH

PANELES AAP LINDOSA		REFERENCIA DEL METADATO			
		PROCESAMIENTO	RESOLUCIÓN		FECHA
PANEL DANTAS	DNT_CA	AGISOFT - RECAP	PNG	16K	ago-25
PANEL PRINCIPAL CERRO AZUL	PP_CA	AGISOFT - RECAP	PNG	16K	ago-25
PANEL LARGO CERRO AZUL	PL_CA	AGISOFT - RECAP	PNG	16K	ago-25
PANEL PRINCIPAL NUEVO TOLIMA	PP_NT	AGISOFT - RECAP	PNG	16K	ago-25
PANEL CORRALITO NUEVO TOLIMA	CRLTO_NT	AGISOFT - RECAP	PNG	16K	ago-25
PANEL RAUDAL GUAYABERO	RDL_GY	AGISOFT - RECAP	PNG	16K	ago-25

Por otro lado, también fue suministrada la simbología oficial de clasificación preventiva del Instituto Colombiano de Antropología e Historia (ICANH) en formato PDF, la cual se presenta en el Anexo 1. A partir de estos insumos se realizó el proceso de selección de los deterioros considerados en el desarrollo experimental. El primer criterio empleado correspondió a la facilidad de identificación visual de las regiones afectadas dentro de los mapas de deterioro. Este criterio fue especialmente relevante debido a que el proceso de construcción de las máscaras binarias requería trasladar manualmente las anotaciones presentes en los mapas hacia las imágenes originales. Con el fin de garantizar la consistencia y confiabilidad de este procedimiento, se priorizaron aquellos deterioros cuyas características gráficas y delimitaciones pudieran ser reconocidas de manera clara e inequívoca a partir de la información suministrada por el ICANH. De esta forma, se buscó minimizar posibles errores de interpretación durante la generación de las máscaras utilizadas como referencia para el entrenamiento de los modelos de segmentación semántica.

Un segundo criterio considerado durante la selección de los deterioros fue su presencia en los paneles documentados disponibles para el estudio. Para garantizar la viabilidad del diseño experimental y la aplicación de estrategias de validación cruzada basadas en la exclusión de una imagen completa durante cada iteración (*Leave-One-Image-Out*, LOIO), era necesario que cada tipo de deterioro estuviera representado en la mayoría de los paneles incluidos en el conjunto de datos. En consecuencia, se priorizaron aquellos deterioros presentes en al menos tres de las cuatro imágenes seleccionadas para el desarrollo experimental. Este criterio permitió asegurar la disponibilidad de ejemplos suficientes tanto para el entrenamiento como para la evaluación de los modelos, favoreciendo una valoración más consistente de su capacidad de generalización frente a imágenes no observadas durante el proceso de aprendizaje.

Con base en los criterios de facilidad de identificación visual y representatividad dentro de los paneles documentados, se seleccionaron los siguientes tipos de deterioro para el desarrollo del estudio: plantas superiores, galerías, nidos y nidos de avispas. Estas categorías presentaban características visuales claramente diferenciables dentro de los mapas de deterioro suministrados por el ICANH y contaban con una distribución suficiente en los paneles disponibles, permitiendo la construcción de conjuntos de datos adecuados para el entrenamiento y validación de modelos de segmentación semántica. Asimismo, dichos deterioros corresponden a alteraciones de origen biológico que pueden afectar la conservación de los paneles rupestres, por lo que su identificación y monitoreo resultan de interés dentro de las estrategias de conservación preventiva del patrimonio arqueológico.

5.2 Construcción de máscaras binarias.

El objetivo de esta etapa metodológica fue generar una máscara binaria para cada tipo de deterioro seleccionado, de manera que pudiera utilizarse como referencia durante el entrenamiento supervisado de los modelos de segmentación semántica. Estas máscaras constituyen la representación digital de las regiones afectadas por cada deterioro y funcionan como la verdad de referencia (*ground truth*) empleada para que los modelos aprendan a identificar y delimitar automáticamente dichas áreas en imágenes previamente no observadas.

La construcción de las máscaras binarias representó una de las fases más críticas del proyecto, debido tanto a la complejidad visual de algunos deterioros como a las características de los insumos suministrados por el ICANH. Adicionalmente, el tamaño de las imágenes utilizadas, con resoluciones que alcanzan los 15360×8640 píxeles, incrementó significativamente el esfuerzo requerido para realizar el proceso de anotación y validación de cada una de las regiones de interés.

La Figura 3 resume el procedimiento general empleado para la construcción de las máscaras binarias utilizadas en el proyecto. El proceso parte de los insumos suministrados por el ICANH, incluyendo las imágenes originales y los mapas de deterioro, continúa con la identificación y delimitación manual de las regiones de interés mediante ImageJ, y finaliza con la generación de una representación binaria que puede emplearse como verdad de referencia (*ground truth*) durante el entrenamiento de los modelos de segmentación semántica.



Figura 3. Flujo metodológico para la generación de máscaras binarias a partir de los mapas de deterioro suministrados por el ICANH.

Como ejemplo del procedimiento empleado, se seleccionó el deterioro denominado F.B. Plantas superiores, correspondiente a un factor biológico asociado al crecimiento de vegetación sobre la superficie del panel rupestre. La Figura 4 presenta el deterioro identificado dentro del mapa de deterioro suministrado por el ICANH y utilizado como referencia durante el proceso de construcción de la máscara binaria.

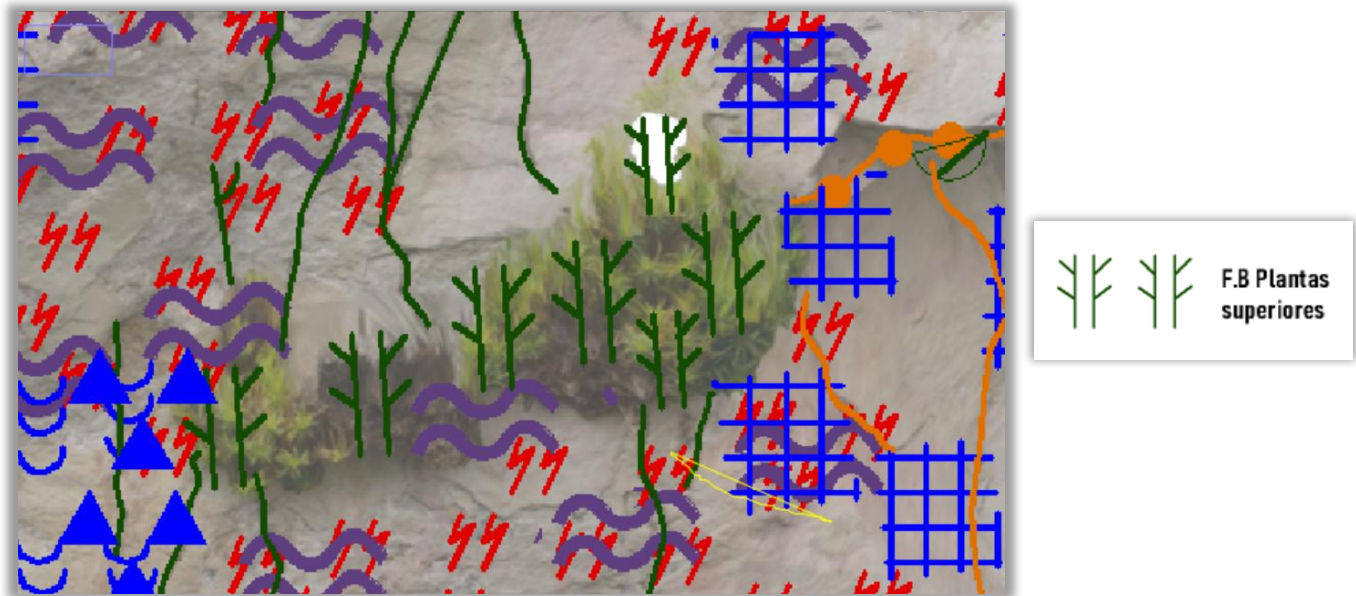


Figura 4. Ejemplo de Factor biológico plantas superiores en mapa de deterioro.

El procedimiento inició con la identificación de la región afectada en el mapa de deterioro y su posterior localización sobre la imagen original del panel. Utilizando el software ImageJ, se delimitó manualmente el contorno de la zona de interés mediante las herramientas *Polygon selections* disponibles en la aplicación. Una vez definida la región correspondiente al deterioro, se aplicó la función *Clear Outside*, eliminando el contenido exterior a la selección y conservando únicamente el área asociada al deterioro

analizado, tal como se observa en la Figura 5.

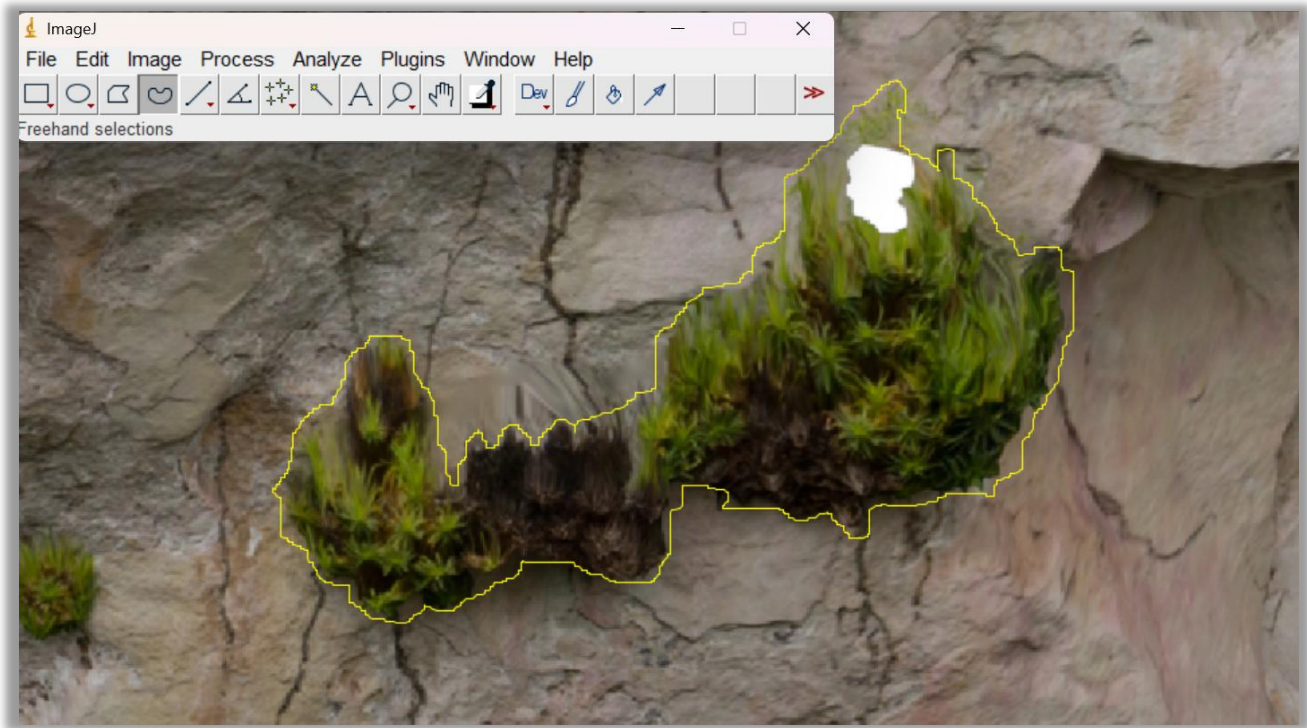


Figura 5. Delimitación del contorno usando polygon selection de ImageJ.

Posteriormente, la imagen resultante fue convertida a escala de grises de 8 bits para facilitar su procesamiento. Sobre esta versión se aplicó la herramienta *Threshold*, permitiendo transformar la región seleccionada en una representación binaria compuesta únicamente por valores de fondo y primer plano. De esta manera, los píxeles pertenecientes al deterioro fueron asignados al valor 255, mientras que el resto de la imagen fue representado mediante el valor 0. El resultado final corresponde a una máscara binaria que puede utilizarse directamente como referencia durante el entrenamiento supervisado de los modelos de segmentación semántica, como se muestra en la

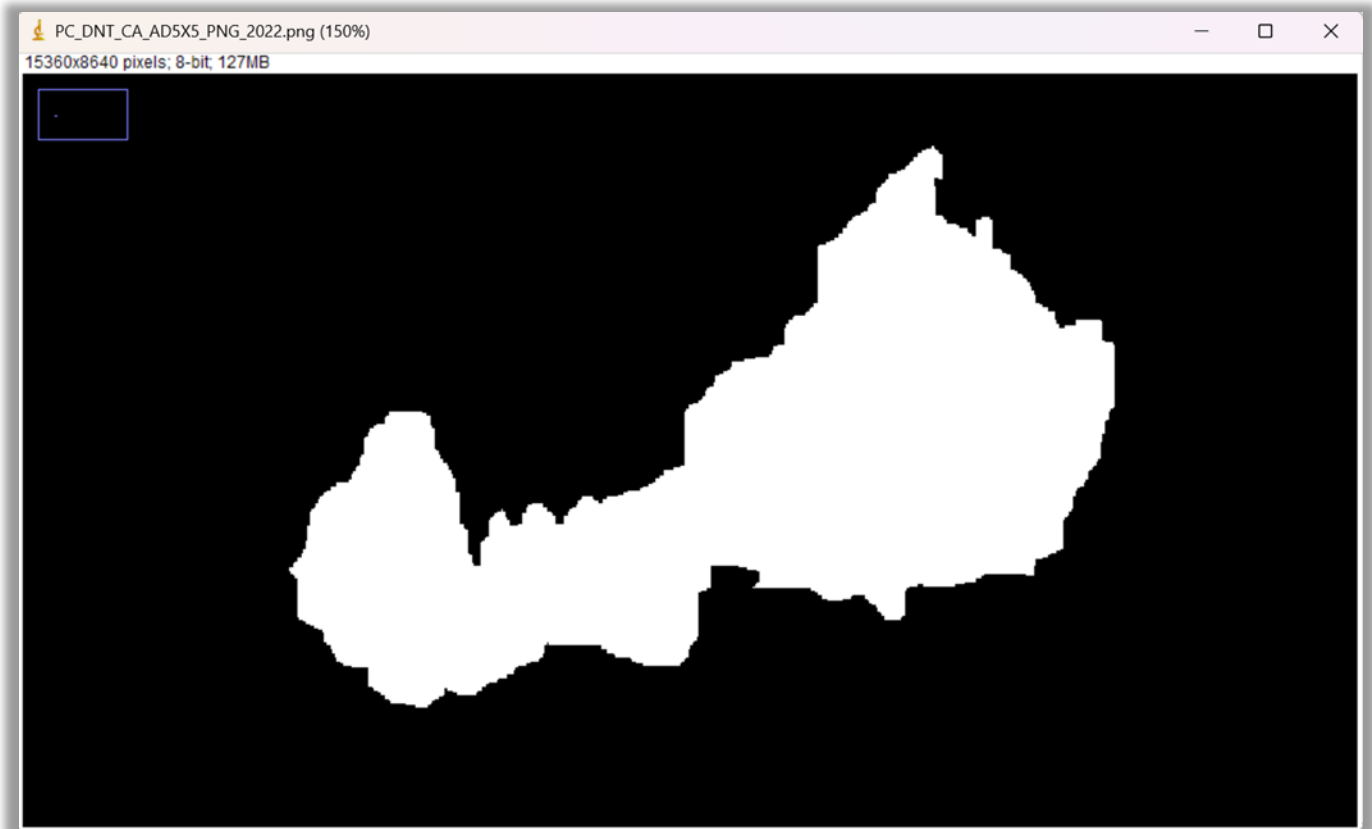


Figura 6. Mascara binaria resultante.

Este proceso fue aplicado de manera sistemática para cada uno de los factores biológicos de deterioro seleccionados y para todos los paneles que contenían ejemplos de dichas categorías. Como resultado, se obtuvo un conjunto de máscaras binarias que representan las regiones afectadas por cada tipo de deterioro y que constituyeron la verdad de referencia (*ground truth*) empleada durante el desarrollo experimental. Estas máscaras fueron utilizadas posteriormente para la extracción de *patches*, la construcción del conjunto de datos y el entrenamiento de los modelos de segmentación semántica descritos en las secciones siguientes.

Con el fin de verificar la calidad de las máscaras binarias generadas y validar la correcta correspondencia espacial entre las regiones deterioradas y las imágenes originales, se realizó una inspección visual utilizando representaciones superpuestas (*overlay*). Las Figura 7, Figura 8 y Figura 9 presentan ejemplos de máscaras binarias obtenidas para el deterioro galerías en los paneles Corralito Nuevo Tolima, Principal Nuevo Tolima y Dantas Cerro Azul, respectivamente. En cada caso se muestra la imagen original, la máscara binaria generada y la superposición de ambas representaciones, permitiendo corroborar que las regiones segmentadas corresponden adecuadamente a las zonas identificadas como deterioro por los especialistas del ICANH.

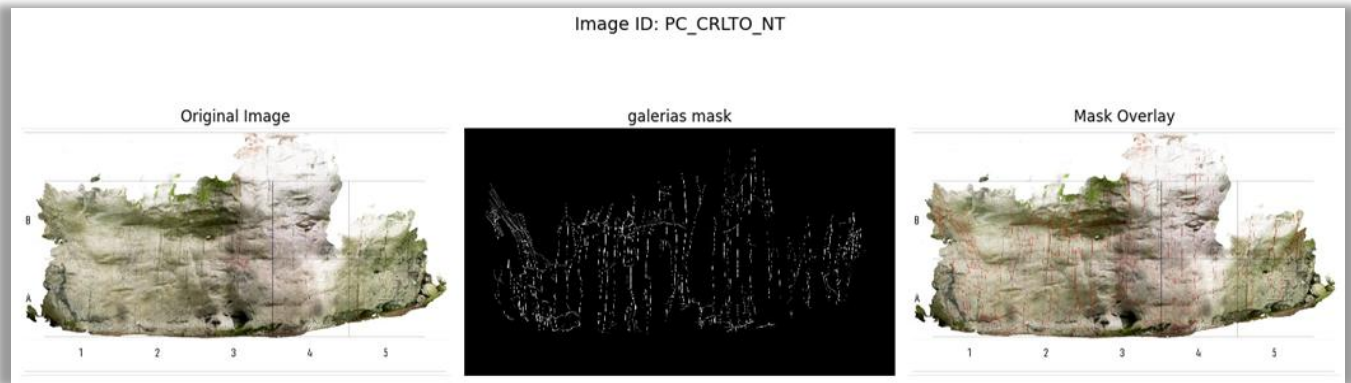


Figura 7. Validación visual de regiones segmentadas correspondientes a deterioros estructurales tipo galerías en el panel PC_CRLTO_NT.

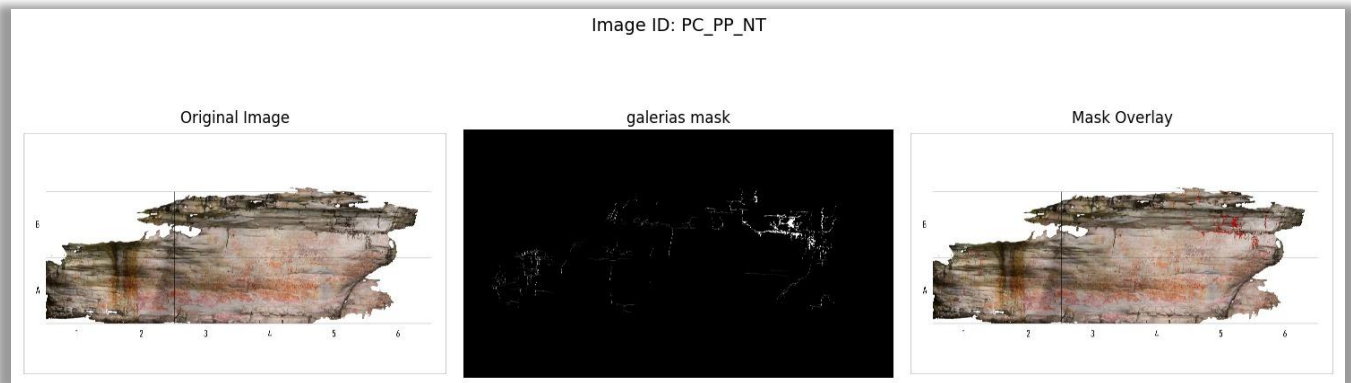


Figura 8. Proceso de generación de máscara binaria para deterioros tipo galerías en el panel PC_PP_NT mediante segmentación manual en ImageJ.

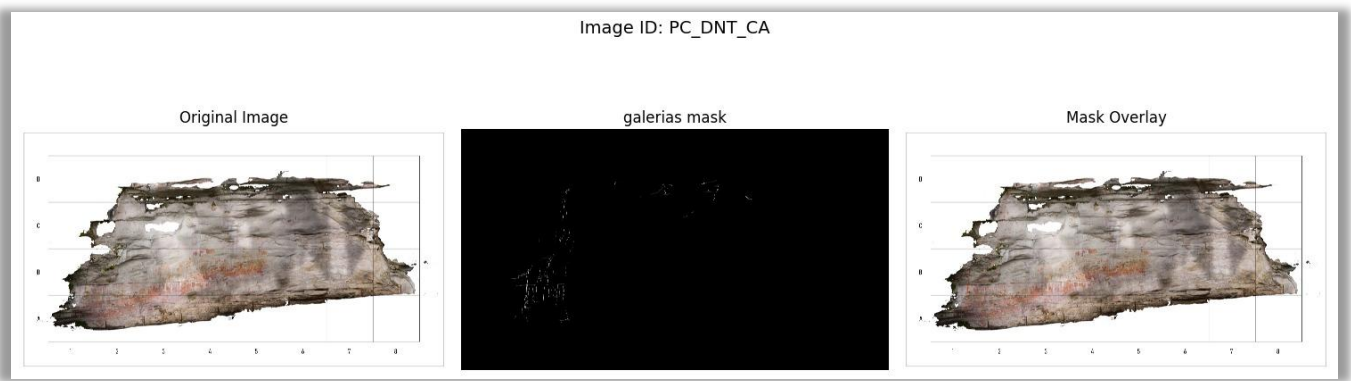


Figura 9. Ejemplo de segmentación manual del deterioro tipo galerías en el panel PC_DNT_CA, incluyendo imagen original, máscara binaria y visualización overlay.

Los resultados obtenidos evidencian que el procedimiento de generación de máscaras permitió preservar adecuadamente la ubicación y morfología de las regiones deterioradas, proporcionando anotaciones consistentes para las etapas posteriores de extracción de *patches* y entrenamiento de los modelos de segmentación semántica.

5.3 Construcción del conjunto de datos para entrenamiento

Como se mencionó en la sección 5.2, las máscaras binarias generadas fueron utilizadas para construir el conjunto de datos empleado durante el entrenamiento y evaluación de los modelos de segmentación semántica. Debido a las características de las imágenes disponibles y a las necesidades del proceso de aprendizaje supervisado, fue necesario implementar varias etapas de preparación y organización de los datos.

5.3.1 Generación de *patches*

Debido a las altas resoluciones de las imágenes suministradas por el ICANH, las cuales alcanzan dimensiones de hasta 15360×8640 píxeles, resultaba inviable utilizar las imágenes completas durante el entrenamiento de los modelos de segmentación semántica debido a las limitaciones de memoria y capacidad de procesamiento. Por esta razón, se implementó una estrategia de partición en *patches* o subimágenes de menor tamaño, permitiendo reducir el costo computacional del entrenamiento y aumentar el número de muestras disponibles para el aprendizaje. Cada *patch* generado conserva la correspondencia espacial entre la imagen original y su respectiva máscara binaria, requisito fundamental para las tareas de segmentación semántica. Adicionalmente, se generó una representación superpuesta (*overlay*) con el fin de facilitar la validación visual de las regiones anotadas y verificar la correcta alineación entre la imagen y la máscara.

La Figura 10 presenta un ejemplo de un *patch* positivo correspondiente al deterioro galerías. En ella se observa el fragmento extraído de la imagen original (a), la máscara binaria asociada (b) y la superposición de ambas representaciones (c), donde el área resaltada en rojo indica la presencia del deterioro. Este tipo de información constituyó la base del conjunto de datos utilizado durante las fases de entrenamiento y validación de los modelos de segmentación semántica.

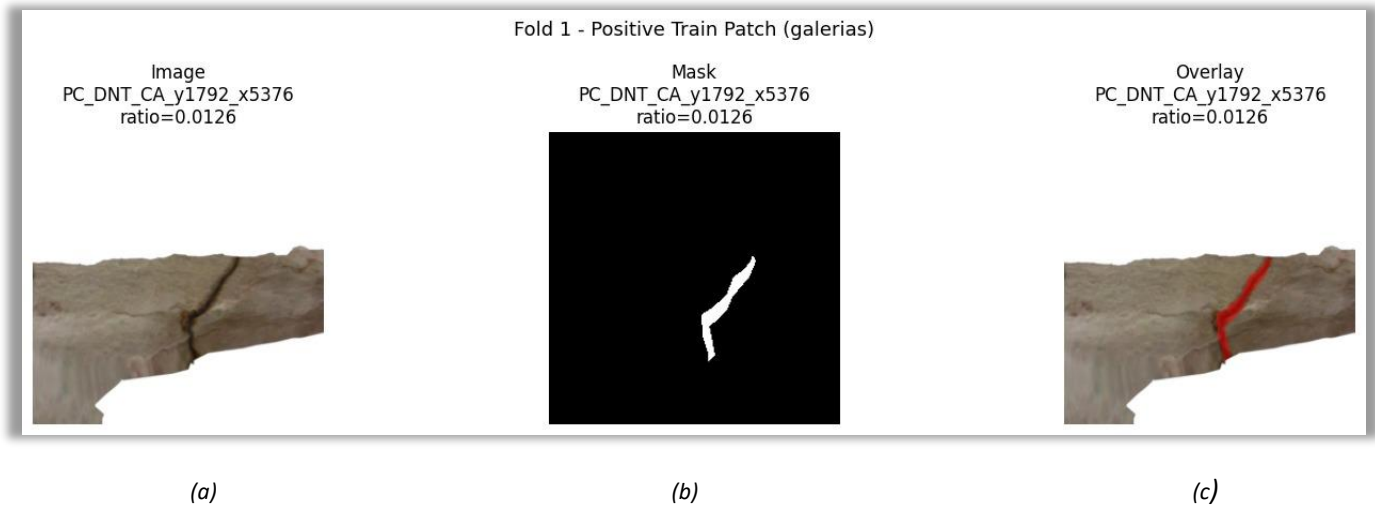


Figura 10. Ejemplo de patch positivo generado para entrenamiento del modelo de segmentación semántica, incluyendo imagen original (a), máscara binaria (b) y visualización overlay del deterioro identificado (c).

5.3.2 Parámetros para la generación de patches

Considerando las dimensiones de las imágenes originales suministradas por el ICANH, se estableció inicialmente un tamaño de *patch* de 256×256 píxeles. Esta configuración permitió dividir cada panel en aproximadamente 1 980 subimágenes, proporcionando un equilibrio adecuado entre el nivel de detalle conservado y los requerimientos computacionales asociados al entrenamiento de los modelos de segmentación semántica. Durante el proceso de extracción, cada *patch* fue generado de forma simultánea a partir de la imagen original y su correspondiente máscara binaria, preservando la correspondencia espacial necesaria para las tareas de segmentación semántica. Posteriormente, se calculó la proporción de píxeles pertenecientes al deterioro dentro de cada subimagen mediante el análisis de la máscara asociada.

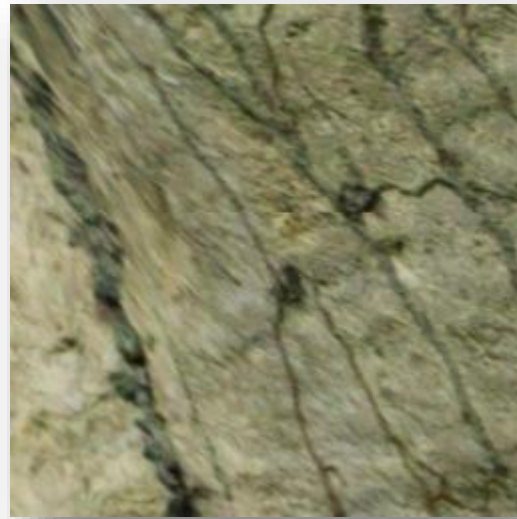
Esta métrica permitió cuantificar el porcentaje de afectación presente en cada muestra y clasificar los *patches* como positivos o negativos, según la presencia o ausencia de regiones deterioradas. La información obtenida resultó fundamental para caracterizar la distribución de las muestras disponibles en el conjunto de datos, identificar posibles problemas de desbalance entre clases y apoyar las estrategias de selección y balanceo implementadas durante las etapas posteriores del desarrollo experimental. En cada *patch* se calcula la proporción de píxeles positivos asociados al deterioro presente, permitiendo seleccionar regiones con información relevante para el entrenamiento supervisado del modelo.

El caso particular del deterioro galerías requirió un análisis adicional debido a las características morfológicas de las regiones afectadas. Como puede observarse en las Figura 7, Figura 8 y Figura 9 este tipo de deterioro se manifiesta principalmente mediante estructuras lineales y de reducido espesor, ocupando una proporción relativamente pequeña del área total de los paneles. En consecuencia, al utilizar *patches* de 256×256 píxeles, la cantidad de píxeles asociados al deterioro dentro de cada subimagen tendía a ser muy baja, generando muestras dominadas por regiones de fondo y con escasa información relevante para el aprendizaje.

Con el propósito de incrementar la representatividad de las zonas afectadas y mejorar la capacidad del modelo para capturar las características visuales y morfológicas de este tipo de deterioro, se optó por reducir el tamaño de los *patches* a 128×128 píxeles durante los experimentos realizados para la categoría galerías como los que se evidencian en la Figura 11. Esta modificación permitió aumentar la densidad relativa de píxeles pertenecientes al deterioro dentro de cada muestra, así como generar un mayor número de ejemplos relevantes para el entrenamiento. Como resultado, se obtuvo un conjunto de datos más adecuado para el aprendizaje de patrones asociados a este tipo de alteración.



(a)



(b)

Figura 11. Detalle de patch resultante con dimensión 128×128 para galerías.

5.3.3 Organización y trazabilidad del conjunto de datos

Con el propósito de garantizar la reproducibilidad de los experimentos y facilitar la gestión de la información generada durante el proceso de entrenamiento y validación, se diseñó una estructura organizada para el almacenamiento de imágenes originales, máscaras binarias, patches y metadatos asociados. La estructura del conjunto de datos fue organizada jerárquicamente por tipo de deterioro, folds experimentales y subconjuntos de entrenamiento y prueba. Adicionalmente, se implementaron archivos de metadatos en formato CSV encargados de almacenar la información descriptiva de cada patch generado, incluyendo su ubicación espacial dentro de la imagen original, las rutas de almacenamiento de la imagen y la máscara, la proporción de deterioro presente y la asignación experimental correspondiente.

El archivo principal metadata.csv almacena la información general de las imágenes originales y sus máscaras asociadas, mientras que los archivos derivados por cada tipo de deterioro contienen los metadatos específicos de los patches utilizados durante el desarrollo experimental. La presenta una descripción de los principales campos registrados en dichos archivos.

Tabla 2. Descripción de los principales campos utilizados en los archivos de metadatos de los patches generados para entrenamiento y validación.

Campo	Descripción
patch_id	Identificador único del <i>patch</i>
source_image_id	Identificador de la imagen original de la cual fue extraído el <i>patch</i>
fold_id	Fold experimental asignado para validación
split	Conjunto al que pertenece el <i>patch</i> (train o test)
target_key	Tipo de deterioro asociado al <i>patch</i>
image_patch_path	Ruta de almacenamiento de la subimagen
mask_patch_path	Ruta de almacenamiento de la máscara binaria
ratio_deterioration	Proporción de píxeles correspondientes al deterioro dentro del <i>patch</i>

5.3.4 Estrategia de partición y validación experimental

La estrategia de validación implementada se basó en el esquema Leave-One-Image-Out (LOIO), utilizando los cuatro paneles disponibles para cada tipo de deterioro. Este enfoque consiste en reservar una imagen completa para evaluación mientras las imágenes restantes son utilizadas durante el proceso de entrenamiento. El procedimiento se repite de manera iterativa hasta que cada imagen ha sido utilizada una vez como conjunto de prueba.

A diferencia de una partición aleatoria de patches, la estrategia LOIO garantiza que todos los patches derivados de una misma imagen permanezcan agrupados dentro del mismo conjunto experimental. De esta manera, se evita que regiones provenientes de un mismo panel aparezcan simultáneamente en entrenamiento y prueba, reduciendo el riesgo de fuga de información (data leakage) y permitiendo una evaluación más realista de la capacidad de generalización de los modelos frente a imágenes no observadas previamente.

La adopción de esta estrategia respondió a la necesidad de evitar fenómenos de fuga de información (*data leakage*) durante la evaluación de los modelos. Dado que cada imagen original genera un gran número de patches espacialmente relacionados, una división aleatoria de las muestras podría provocar que regiones provenientes de una misma imagen estuvieran presentes simultáneamente en los conjuntos de entrenamiento y prueba. Esta situación podría producir una sobreestimación del desempeño real del modelo, al evaluar ejemplos con características visuales muy similares a las observadas durante el entrenamiento.

Para el caso del deterioro galerías, se construyeron cuatro folds experimentales, donde cada uno utilizó tres paneles para entrenamiento y un panel independiente para evaluación. Las métricas reportadas corresponden al promedio obtenido a través de las cuatro iteraciones realizadas.

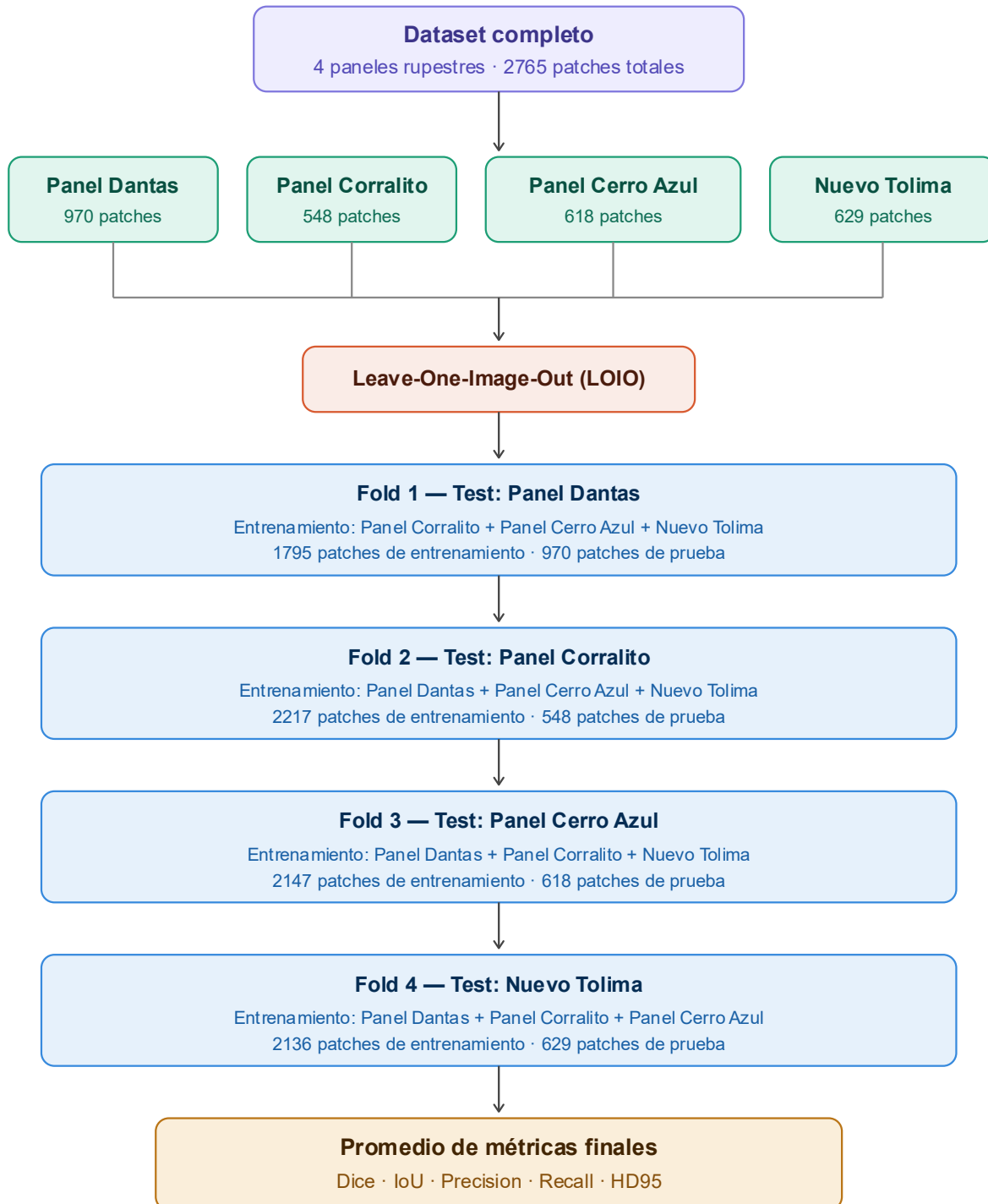


Figura 12. Proceso de validación usando LOIO.

La Figura 12 ilustra el esquema de validación Leave-One-Image-Out (LOIO) implementado en el estudio. En cada iteración experimental, un panel completo es reservado para evaluación mientras los paneles restantes son utilizados para entrenamiento. Este procedimiento se repite hasta que cada panel ha sido utilizado una vez como conjunto de prueba, permitiendo obtener una estimación más robusta de la capacidad de generalización de los modelos frente a imágenes no observadas previamente.

Una vez generados los *patches*, se realizó un análisis de la distribución de las muestras positivas y negativas disponibles para cada tipo de deterioro. Debido a que gran parte de la superficie de los paneles corresponde a regiones sin afectación, la cantidad de *patches* negativos resultó considerablemente superior al número de muestras que contenían deterioro. Con el fin de reducir este desbalance y evitar que el modelo favoreciera excesivamente la clase mayoritaria durante el entrenamiento, se implementó una estrategia de submuestreo adaptativo (*adaptive undersampling*). Este procedimiento consistió en conservar la totalidad de los *patches* positivos y seleccionar únicamente una fracción representativa de los *patches* negativos, manteniendo una relación máxima predefinida entre ambas clases. La aplicación de esta estrategia permitió construir conjuntos de entrenamiento más equilibrados, favoreciendo el aprendizaje de patrones asociados a los deterioros y reduciendo el riesgo de que el modelo convergiera hacia soluciones dominadas por la clasificación de fondo

DISEÑO E IMPLEMENTACIÓN DEL MODELO

6.1 Flujo de trabajo para entrenamiento del modelo

Con el objetivo de garantizar flexibilidad experimental y reutilización de componentes, se diseñó un flujo de trabajo modular dividido en diferentes etapas independientes. Esta decisión metodológica permitió separar los procesos de generación de datos, entrenamiento y evaluación, evitando repetir tareas computacionalmente costosas durante las pruebas experimentales y facilitando la comparación entre diferentes arquitecturas de segmentación semántica.

Inicialmente, el flujo de trabajo fue organizado mediante notebooks independientes encargados de la generación de *patches* y creación de la meta data, el entrenamiento de modelos y la evaluación de resultados. Esta estructura permitió ejecutar experimentos de manera más controlada y mantener una separación clara entre las diferentes etapas del desarrollo experimental. Asimismo, el flujo de trabajo fue desarrollado utilizando Python junto con TensorFlow y Keras como frameworks principales de entrenamiento, manteniendo compatibilidad parcial con PyTorch para futuras extensiones y pruebas comparativas.

De manera general, el proceso implementado incluyó las siguientes etapas:

- generación automática de patches;
- carga optimizada de imágenes;
- aumento de datos;
- entrenamiento del modelo;
- validación experimental;
- evaluación de métricas;
- exportación de resultados.

6.2 Arquitecturas evaluadas

En la etapa inicial de investigación se llevó a cabo una revisión comparativa de arquitecturas ampliamente empleadas en tareas de segmentación semántica. La selección preliminar se orientó hacia modelos con desempeño destacado en escenarios caracterizados por conjuntos de datos de tamaño reducido y regiones de interés morfológicamente complejas, condiciones que corresponden a las presentes en este proyecto. Entre las arquitecturas evaluadas se consideraron *U-Net*, *DeepLabv3+*, *ResUNet*, *Attention U-Net* y diversas variantes híbridas basadas en *ResNet*. No obstante, la fase experimental se concentró en arquitecturas derivadas de *U-Net*, atendiendo al volumen limitado del conjunto de datos disponible, a los recursos computacionales accesibles durante el desarrollo del proyecto y a su consolidada aplicación en problemas de segmentación biomédica y de patrimonio cultural.

Las arquitecturas finalmente implementadas fueron *U-Net* y *ResUNet*, seleccionadas por su capacidad para preservar información espacial a través de conexiones de salto (*skip connections*) y por los resultados favorables que ambas han reportado en tareas de segmentación con conjuntos de datos reducidos. Con el propósito de sistematizar la comparación entre modelos, se diseñó una estructura modular que permite intercambiar dinámicamente la arquitectura activa mediante parámetros de configuración, sin necesidad de modificar el flujo general de entrenamiento. Esta modularidad no solo facilitó la gestión eficiente de imágenes de alta resolución, sino que también favoreció la reutilización de componentes a lo largo de las distintas etapas experimentales.

6.3 Configuración experimental

Con el propósito de garantizar la comparabilidad entre los diferentes experimentos realizados, se definió una configuración base de entrenamiento aplicable a las arquitecturas evaluadas durante el desarrollo del proyecto. Esta configuración incluyó parámetros relacionados con el tamaño de entrada, optimización, criterios de convergencia y estrategias de regularización orientadas a mejorar la estabilidad del proceso de aprendizaje. Las imágenes fueron procesadas mediante *patches* previamente generados a partir de las representaciones originales de los paneles rupestres. Dependiendo de las características morfológicas del deterioro analizado, se emplearon tamaños de patch de 256×256 píxeles o 128×128 píxeles, preservando en todos los casos la correspondencia entre la imagen y su respectiva máscara binaria. Las muestras fueron organizadas siguiendo la estrategia de validación Leave-One-Image-Out (LOIO), garantizando la independencia entre los conjuntos de entrenamiento y prueba.

Para el proceso de optimización se utilizó el algoritmo Adam, ampliamente empleado en tareas de aprendizaje profundo debido a su capacidad para adaptar dinámicamente la tasa de aprendizaje durante el entrenamiento. La tasa de aprendizaje inicial fue establecida en 5×10^{-5} , valor seleccionado tras diferentes pruebas preliminares orientadas a favorecer la estabilidad de la convergencia y reducir fluctuaciones excesivas en la función de pérdida. La función objetivo utilizada combinó la pérdida Binary Cross Entropy (BCE) con la métrica Dice Loss. Esta combinación permitió equilibrar el aprendizaje a nivel de píxel con la capacidad de segmentar correctamente regiones deterioradas de tamaño reducido, aspecto particularmente importante debido al desbalance existente entre las áreas de fondo y las zonas afectadas dentro de los paneles rupestres.

El entrenamiento se ejecutó utilizando un tamaño de lote (*batch size*) de 8 muestras y un máximo de 100 épocas por experimento. Adicionalmente, se implementó un mecanismo de reducción adaptativa de la tasa de aprendizaje mediante la estrategia ReduceLROnPlateau, permitiendo disminuir automáticamente el valor del *learning rate* cuando las métricas de validación permanecían estables durante varias épocas consecutivas. Con el fin de evitar fenómenos de sobreajuste (*overfitting*) y reducir tiempos de entrenamiento innecesarios, se incorporó una estrategia de parada temprana (*Early Stopping*). Este mecanismo monitoreó el comportamiento de las métricas de validación y detuvo automáticamente el entrenamiento cuando no se observaron mejoras significativas durante un número determinado de iteraciones.

Tabla 3. Hiperparámetros iniciales seleccionados para el entrenamiento.

Parámetro	Valor
Framework de entrenamiento	PyTorch
Arquitecturas evaluadas	U-Net, ResUNet
Tamaño de entrada	256×256 px / 128×128 px
Batch size	8
Épocas máximas	100
Learning rate inicial	5×10^{-5}
Optimizador	Adam
Función de pérdida	BCE + Dice Loss
Scheduler	ReduceLROnPlateau
Early Stopping	Sí
Estrategia de validación	Leave-One-Image-Out (LOIO)

La Tabla 3 resume los principales hiperparámetros utilizados durante los experimentos realizados

RESULTADOS

En este capítulo se presentan los resultados obtenidos durante la construcción del conjunto de datos, el entrenamiento de los modelos de segmentación semántica y la evaluación experimental realizada sobre los diferentes tipos de deterioro seleccionados. Los experimentos fueron desarrollados siguiendo la metodología descrita en los capítulos anteriores, utilizando máscaras binarias generadas a partir de los mapas de deterioro suministrados por el ICANH y aplicando una estrategia de validación *Leave-One-Image-Out* (LOIO) para garantizar la independencia entre los conjuntos de entrenamiento y prueba. Los resultados se organizan de acuerdo con cada categoría de deterioro analizada, permitiendo evaluar el comportamiento de los modelos frente a diferentes características morfológicas y niveles de complejidad presentes en los paneles rupestres. Para la evaluación del desempeño se emplearon métricas de segmentación ampliamente utilizadas en la literatura, incluyendo *Dice*, *IoU*, *precisión*, *recall* y *HD95*, con el propósito de obtener una valoración integral de la calidad de las segmentaciones generadas.

7.1 Resultados obtenidos para el deterioro plantas superiores

Durante la fase inicial de experimentación se entrenó un modelo de segmentación semántica basado en la arquitectura U-Net, utilizando *patches* de 256×256 píxeles correspondientes a la clase plantas superiores. Esta clase fue seleccionada como punto de partida debido a que contaba con un conjunto de máscaras binarias más consistente frente a otras categorías de deterioro. Inicialmente se decidió hacer el entrenamiento con la totalidad de los *patches* disponibles, esto con el fin de evaluar experimentalmente el impacto del balanceo de clases en el entrenamiento del modelo.

7.1.1 Evaluación del modelo U-Net sin balanceo de clases

El modelo implementado corresponde a una arquitectura U-Net con bloques convolucionales, normalización por lotes (*Batch Normalization*), capas de regularización mediante *Dropout* y conexiones de salto (*skip connections*), lo que permite conservar información espacial relevante durante el proceso de segmentación. En esta primera aproximación, el entrenamiento se llevó a cabo con una estrategia de *early stopping*, deteniéndose en la época 31, lo cual sugiere una convergencia relativamente rápida bajo la configuración experimental utilizada.

Como puede observarse en la

Figura 13, las métricas Dice e IoU presentan una distribución bimodal. La mayor parte de los *patches* obtiene valores cercanos a 1, indicando una elevada coincidencia entre las máscaras predichas y las máscaras de referencia. Sin embargo, también se observa un grupo de muestras con valores próximos a cero, correspondiente a casos donde el modelo no logró segmentar correctamente las regiones deterioradas o donde la presencia del deterioro era muy reducida dentro del patch. En contraste, la distribución de la métrica HD95 evidencia que la mayoría de las muestras presentan distancias pequeñas entre los contornos predichos y los reales, mientras que un número reducido de patches genera valores considerablemente superiores, los cuales aparecen como valores atípicos en el diagrama de caja. Estos casos corresponden

generalmente a errores de segmentación localizados o a regiones con límites complejos, lo que incrementa significativamente la distancia entre ambas máscaras.

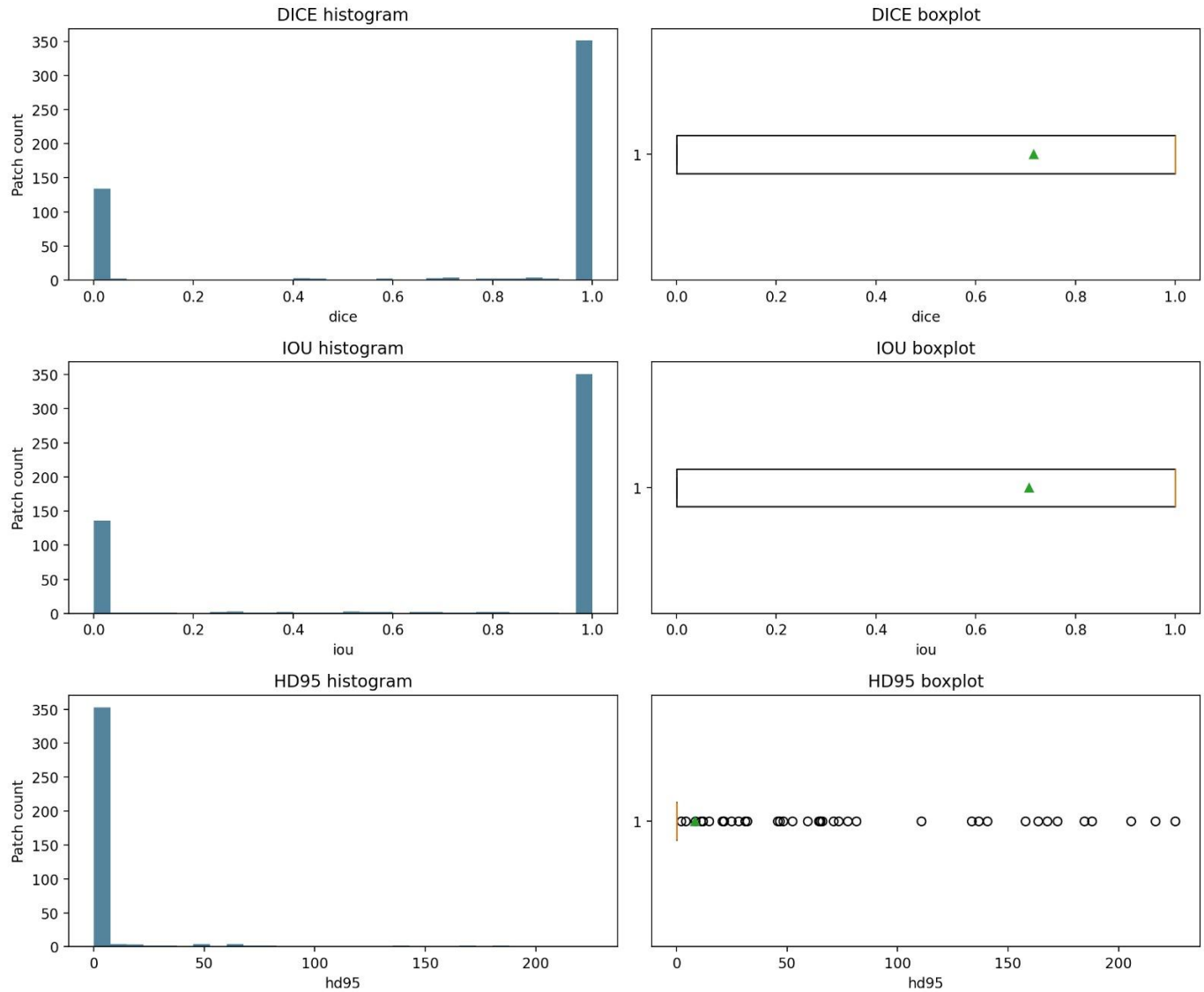


Figura 13. Distribución de métricas Dice, IoU y HD95 obtenidas durante la evaluación del modelo U-Net.

A nivel general, las métricas obtenidas durante la evaluación mostraron valores elevados en una cantidad importante de parches, especialmente en *Dice* e *IoU*. Como se observa en la Figura 13, la distribución de estas métricas presenta una concentración considerable de valores cercanos a 1.0, lo que inicialmente podría interpretarse como un desempeño favorable del modelo. Sin embargo, al analizar con mayor detalle los resultados, se identificó que este comportamiento está fuertemente influenciado por la presencia de una gran cantidad de parches sin deterioro, es decir, parches cuya máscara real se encuentra completamente vacía. En estos casos, el modelo puede obtener valores perfectos simplemente al predecir correctamente la ausencia de la clase objetivo, sin que esto implique necesariamente una buena capacidad para detectar regiones reales de deterioro.

Con el propósito de comprender mejor este comportamiento, se realizó un análisis del desbalance presente en el conjunto de prueba correspondiente al *fold* 1. Los resultados mostraron que, de un total de 523 parches evaluados, 484 parches, equivalentes al 92.54 %, no contenían píxeles positivos asociados a la clase *plantas_superiores*, mientras que únicamente 39 parches, equivalentes al 7.46 %, presentaban deterioro. Esta distribución confirma un desbalance significativo entre clases, donde los parches negativos predominan ampliamente sobre los positivos. Como consecuencia, las métricas globales pueden verse sobreestimadas, debido a que el modelo obtiene buenos resultados en casos triviales donde no existe deterioro que segmentar.

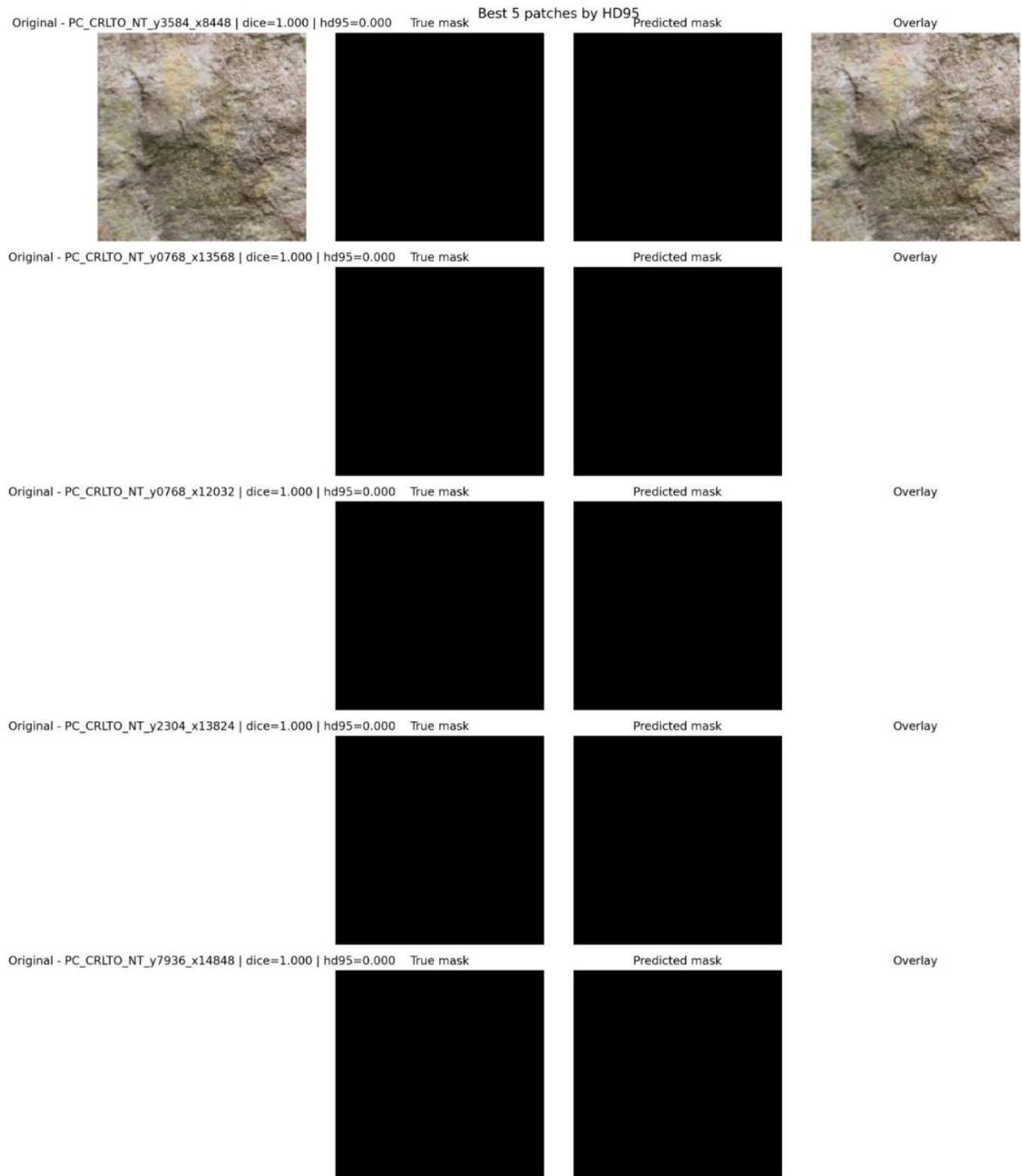


Figura 14. Ejemplos de mejores predicciones del modelo U-Net según la métrica HD95 para la clase plantas superiores.



Figura 15. Ejemplos de mejores predicciones del modelo U-Net según la métrica Dice para la clase plantas superiores.

Las Figura 14 Figura 15 presenta ejemplos individuales de evaluación por *patch*, que tienen las mejores 5 clasificaciones según *HD95* y *Dice* respectivamente. incluyendo la imagen original, la máscara real, la máscara predicha y la visualización superpuesta. Este tipo de representación permitió realizar una validación visual más clara del comportamiento del modelo, ya que no solo muestra el valor numérico de las métricas, sino también la ubicación espacial de los aciertos y errores. En los casos particulares de las figuras antes mencionadas, se puede evidenciar claramente que los mejores casos predichos corresponden en su totalidad a *patches* que no tienen deterioro en ellos, lo que confirma la necesidad de emplear una estrategia efectiva de balanceo. En algunos casos particulares, como los de la Figura 16, se evidencia una correspondencia aceptable entre la máscara real y la predicción; sin embargo, también se observan diferencias en los bordes y pequeñas regiones donde el modelo amplía o reduce la zona segmentada. Este análisis visual fue importante para complementar la interpretación de *Dice* e *IoU*, dado que estas métricas no siempre reflejan completamente los errores espaciales de la predicción.

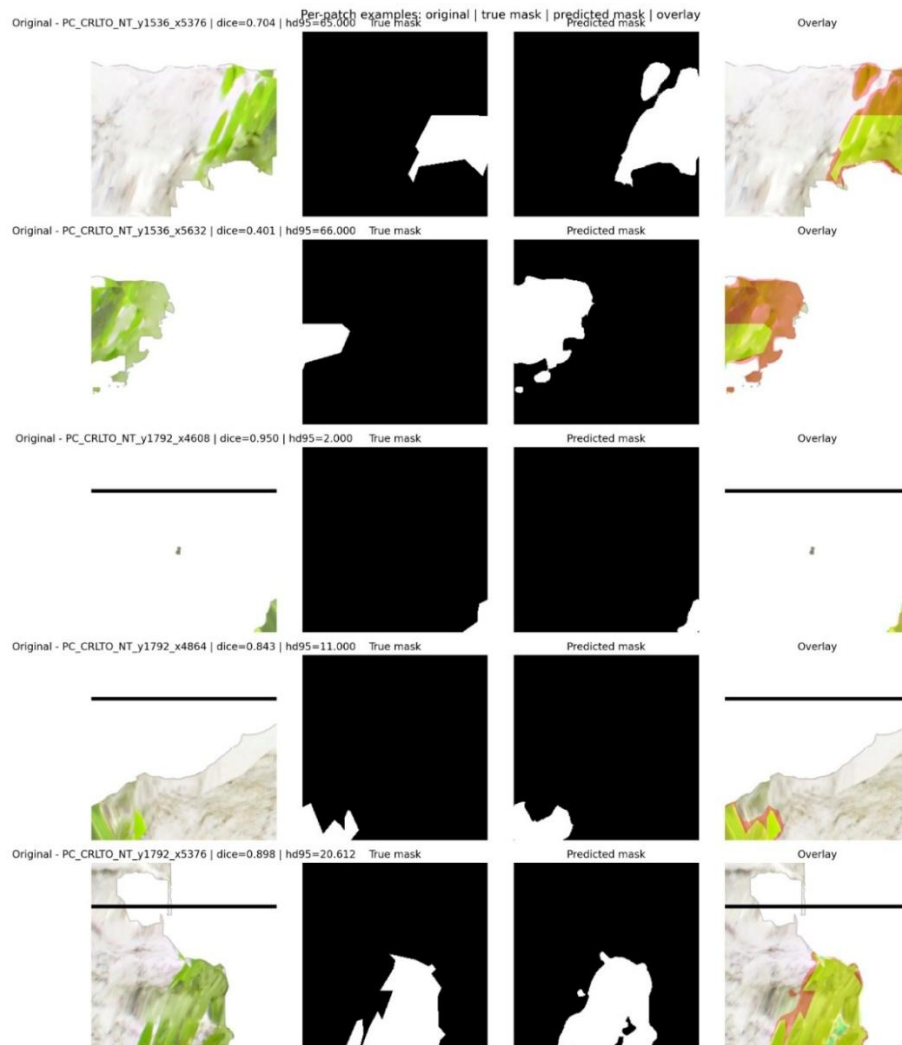


Figura 16. Ejemplos de predicciones por parche del modelo U-Net, incluyendo imagen original, máscara real, máscara predicha y superposición.

Por otra parte, al analizar los peores casos según *HD95* y *Dice*, presentados en las Figura 17 y Figura 18, se identificaron errores relevantes en la segmentación. En varios parches el modelo generó falsos positivos significativos, clasificando regiones de roca, sombra o textura superficial como si correspondieran a plantas superiores. También se observaron casos en los que la máscara real contenía una región pequeña de deterioro, pero la predicción generó una zona mucho más amplia o espacialmente desplazada. Este comportamiento evidencia que el modelo aún presenta dificultades para diferenciar adecuadamente la clase objetivo frente a elementos visuales similares del entorno, especialmente cuando existen variaciones de textura, iluminación o color en la superficie rocosa.

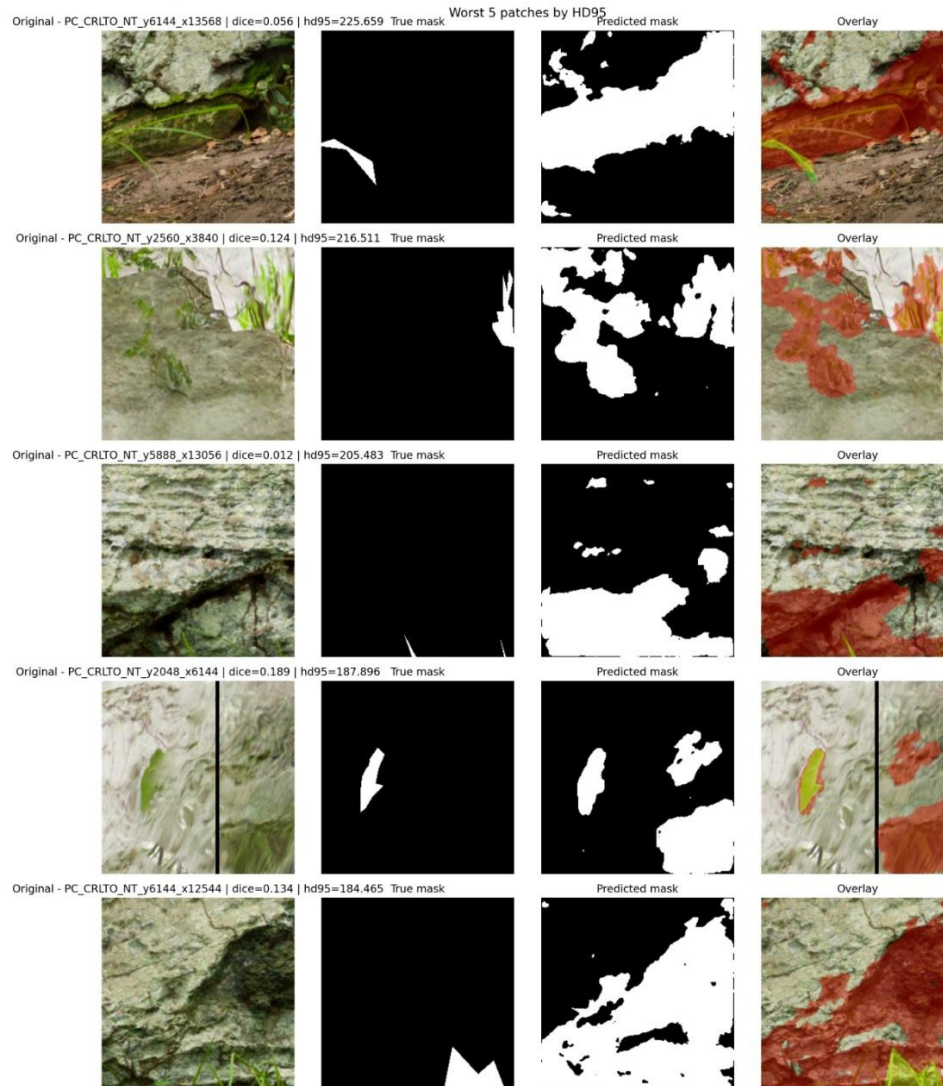


Figura 17. Ejemplos de peores predicciones del modelo U-Net según la métrica *HD95* para la clase plantas superiores.

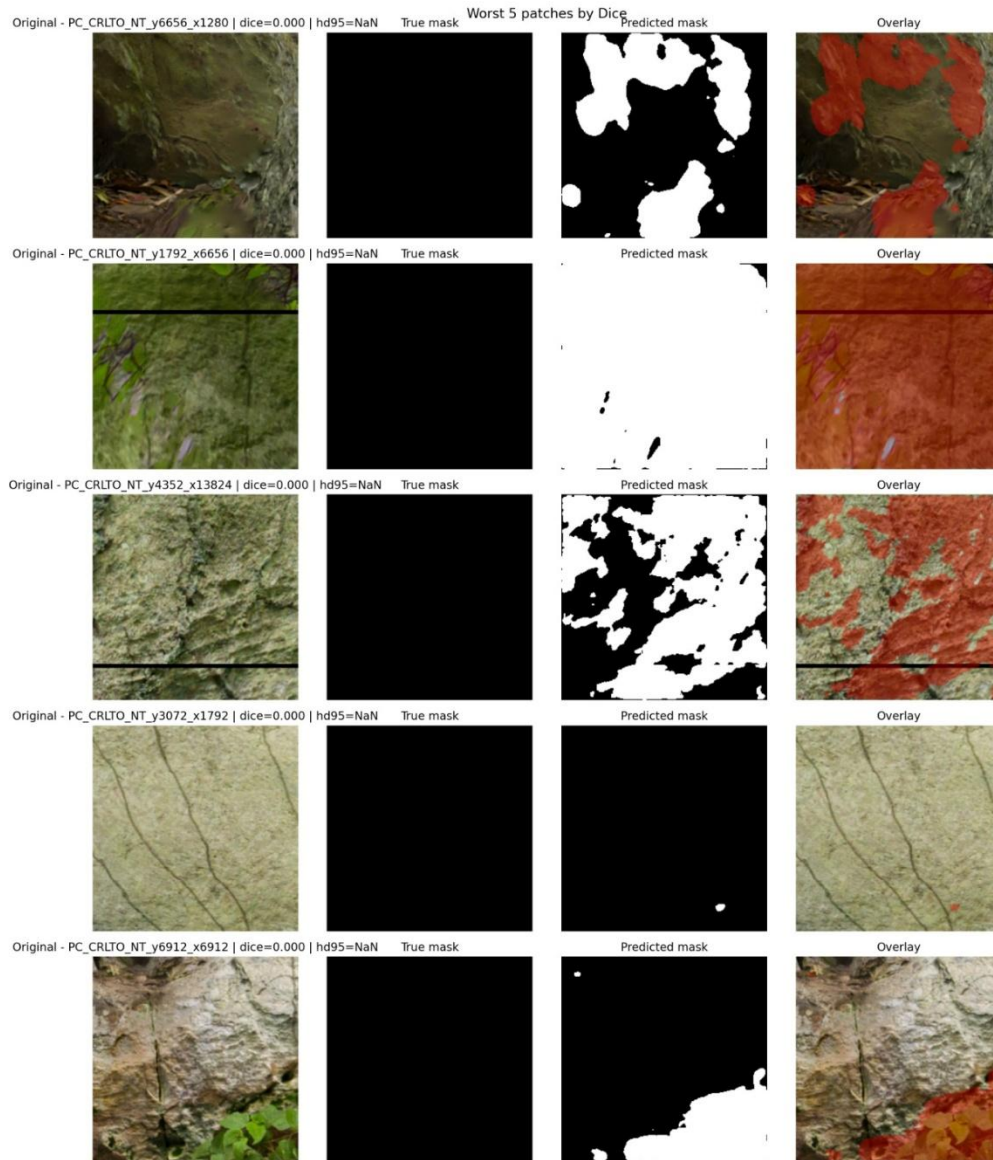


Figura 18. Ejemplos de peores predicciones del modelo U-Net según la métrica Dice para la clase plantas superiores

La incorporación de la métrica *HD95* resultó especialmente útil para identificar errores espaciales que no siempre son evidentes al observar únicamente *Dice* e *IoU*. Mientras estas últimas métricas se enfocan en el nivel de solapamiento entre la máscara real y la predicción, *HD95* permite evaluar diferencias relacionadas con la distancia entre bordes y la localización de las regiones segmentadas. Por esta razón, algunos parches con valores moderados de *Dice* presentaron valores altos de *HD95*, indicando que, aunque existía cierto solapamiento, la predicción no estaba correctamente ubicada o presentaba errores importantes en sus límites. Esto confirma la necesidad de utilizar métricas complementarias en problemas de segmentación semántica, especialmente cuando las anotaciones manuales pueden presentar pequeñas imprecisiones o cuando la clase objetivo ocupa una proporción reducida del parche.

En síntesis, los resultados preliminares muestran que el modelo U-Net logra aprender adecuadamente la

representación del fondo y resolver correctamente numerosos casos simples; sin embargo, su desempeño en la detección efectiva de regiones con deterioro aún es limitado. El principal factor identificado fue el desbalance del conjunto de datos, dado que la mayoría de los parches evaluados no contenían presencia de la clase objetivo. Estos hallazgos permiten concluir que las métricas globales deben interpretarse con precaución y que es necesario complementar la evaluación con análisis específicos sobre parches positivos. A partir de estos resultados, se plantea como siguiente paso ajustar la estrategia de entrenamiento mediante técnicas de balanceo de datos, como undersampling u otras alternativas, además de explorar funciones de pérdida más adecuadas para clases desbalanceadas.

7.1.2 Evaluación del modelo U-Net con estrategia de undersampling adaptativo

Con el propósito de mitigar el impacto del desbalance identificado en la primera etapa experimental, se implementó una estrategia de balanceo basada en *undersampling* adaptativo sobre el conjunto de entrenamiento. Esta decisión surgió a partir del análisis previo, donde se evidenció que una alta proporción de parches no contenía píxeles positivos asociados a la clase plantas superiores, lo cual generaba una sobreestimación de las métricas globales y favorecía que el modelo aprendiera principalmente la representación del fondo. La estrategia aplicada consistió en reducir la cantidad de parches negativos, es decir, aquellos sin presencia de deterioro, manteniendo como criterio un máximo de tres parches negativos por cada parche positivo. Es importante precisar que este balanceo se realizó únicamente sobre el conjunto de entrenamiento, conservando la distribución original en los conjuntos de validación y prueba, con el fin de evaluar el modelo en condiciones más cercanas al comportamiento real de los datos.

Después de aplicar esta estrategia, se realizó un nuevo entrenamiento del modelo *U-Net* bajo condiciones experimentales similares a las utilizadas en la primera etapa. A diferencia del escenario sin balanceo, las métricas globales presentaron una disminución considerable, especialmente en *Dice* e *IoU*, alcanzando valores promedio cercanos a 0.25. Aunque inicialmente este resultado podría interpretarse como una reducción del desempeño, en realidad representa una medición más exigente y realista del comportamiento del modelo, ya que disminuye el efecto artificial producido por la gran cantidad de parches negativos correctamente clasificados como fondo. En otras palabras, el modelo deja de beneficiarse principalmente de casos triviales y empieza a ser evaluado de forma más directa frente a la dificultad real del problema: identificar correctamente regiones donde sí existe presencia de deterioro.

La distribución de las métricas obtenidas después del undersampling se presenta en la Fig. 17. En comparación con la etapa anterior, se observa una mayor dispersión de los valores de Dice, IoU y HD95, lo cual indica que el modelo ya no se comporta únicamente como un clasificador de fondo, sino que empieza a enfrentarse de manera más activa a los parches donde existe presencia de la clase objetivo. Sin embargo, esta dispersión también evidencia que el desempeño aún es inestable, especialmente en los casos donde la región deteriorada es pequeña, poco contrastada o visualmente similar a otras texturas presentes en la roca. Este comportamiento confirma que el balanceo ayuda a reducir el sesgo hacia la clase mayoritaria, pero no resuelve completamente las dificultades propias de la segmentación.

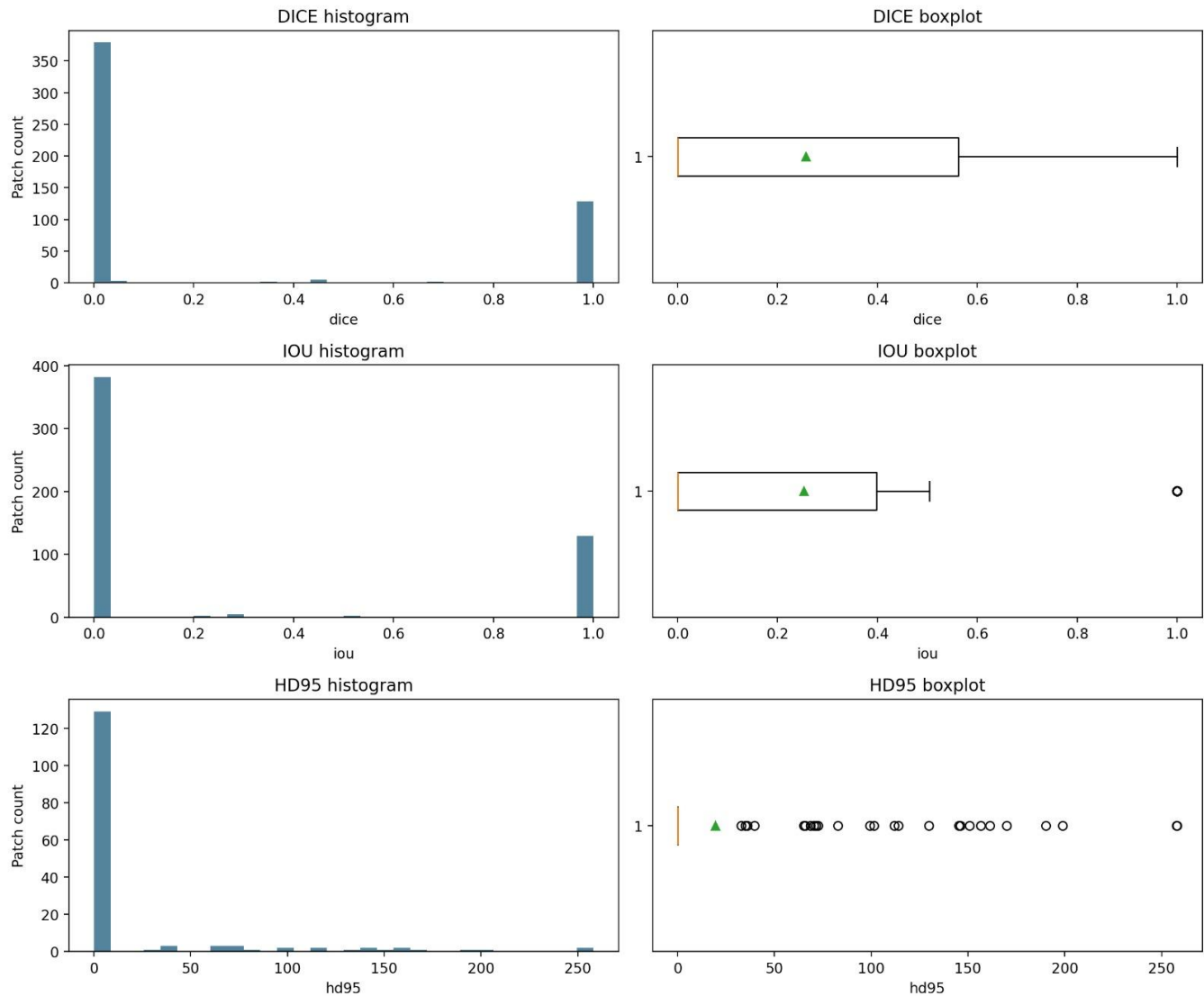


Figura 19. Histogramas y diagramas de caja de las métricas Dice, IoU y HD95 obtenidas durante la evaluación del modelo U-Net con undersampling adaptativo.

El análisis cualitativo de las predicciones permitió identificar un cambio importante en el comportamiento del modelo. Mientras que en la primera etapa el modelo tendía a predecir mayoritariamente fondo, en esta nueva configuración se observa que comienza a reconocer regiones asociadas a la clase plantas superiores. En la Fig. 15 se presentan ejemplos por parche donde se comparan la imagen original, la máscara real, la máscara predicha y la superposición. Estas visualizaciones muestran que el modelo empieza a generar respuestas frente a zonas con presencia de vegetación o deterioro biológico; no obstante, también se evidencian errores de localización, ausencia de detección en algunos parches positivos y predicciones parciales que no alcanzan a cubrir adecuadamente la región anotada manualmente.

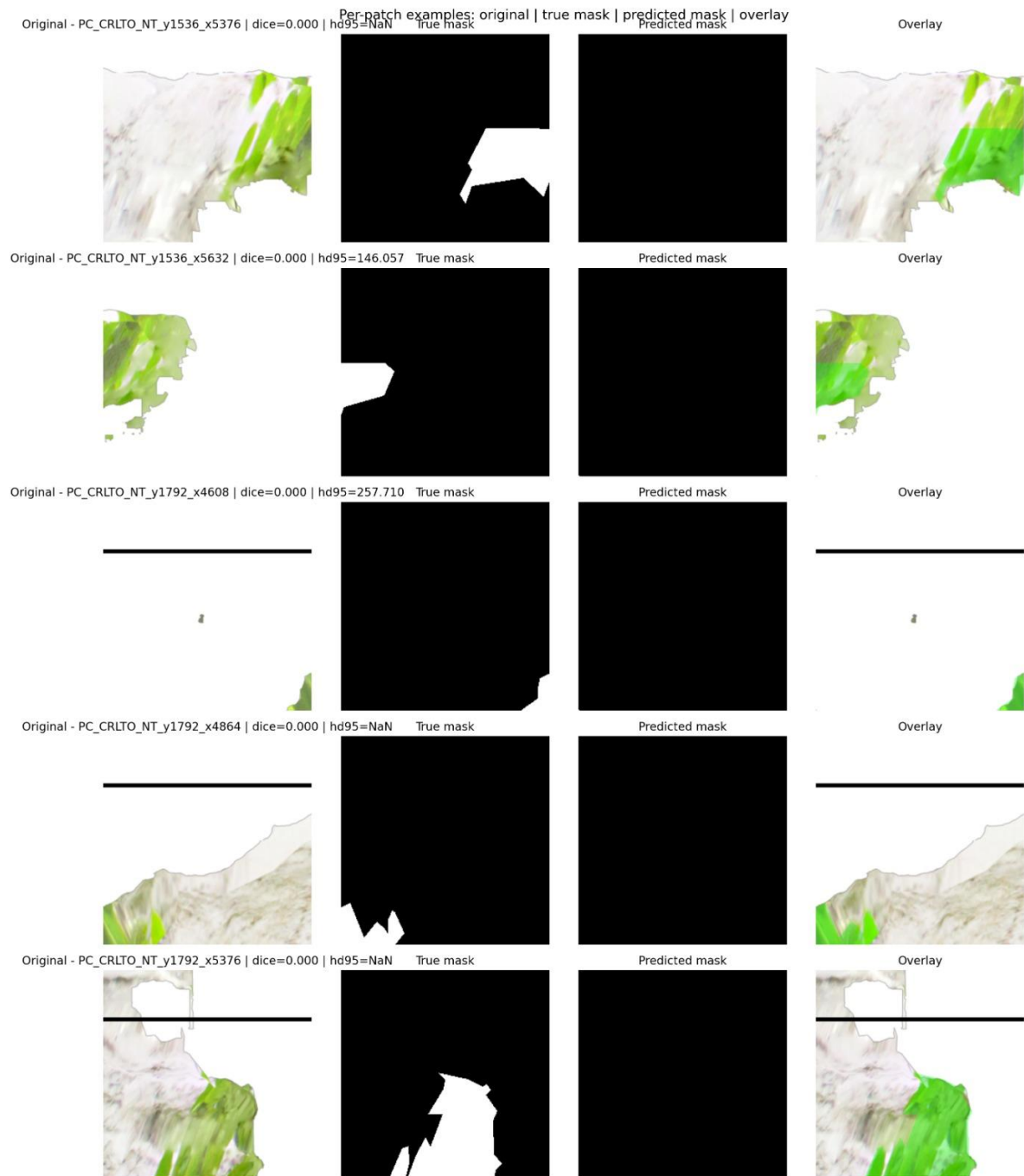


Figura 20. Ejemplos de predicciones por parche del modelo U-Net con undersampling adaptativo, incluyendo imagen original, máscara real, máscara predicha y superposición.

Los mejores resultados obtenidos bajo esta configuración se muestran en la Fig. 16 y la Fig. 17. En estos casos, el modelo logra clasificar correctamente algunos parches sin presencia de deterioro, manteniendo máscaras predichas vacías cuando la máscara real también se encuentra vacía.

Aunque estos resultados reflejan una coincidencia perfecta en términos de Dice o HD95, deben interpretarse con cuidado, ya que muchos corresponden nuevamente a parches negativos. Por esta razón, estos ejemplos son útiles para verificar que el modelo conserva la capacidad de reconocer fondo, pero no necesariamente demuestran una mejora sustancial en la detección de la clase positiva. Esta observación refuerza la necesidad de evaluar de manera diferenciada los parches positivos y negativos en futuras etapas experimentales.

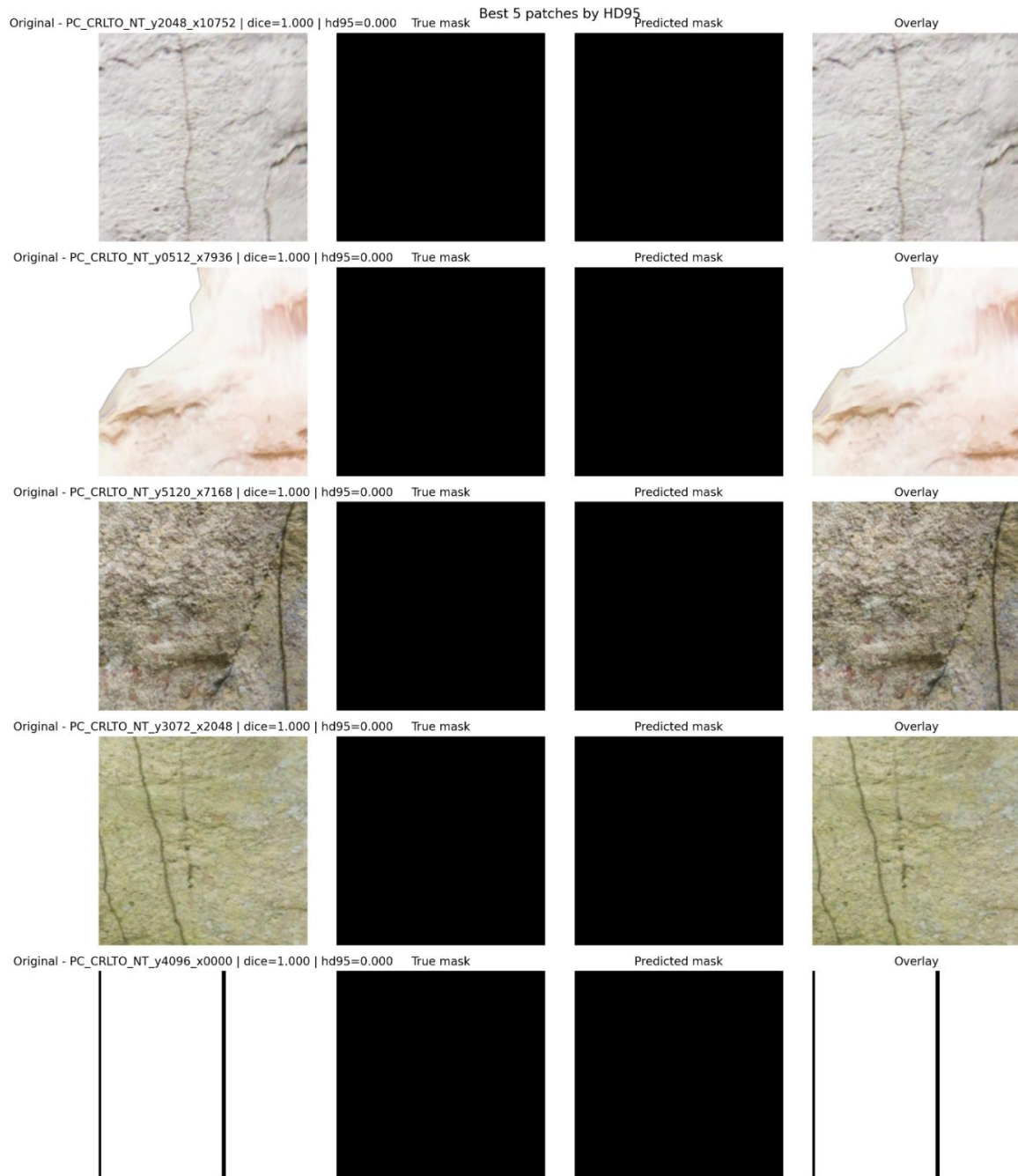


Figura 21. Ejemplos de mejores predicciones del modelo U-Net con undersampling adaptativo según la métrica HD95 para la clase plantas superiores.

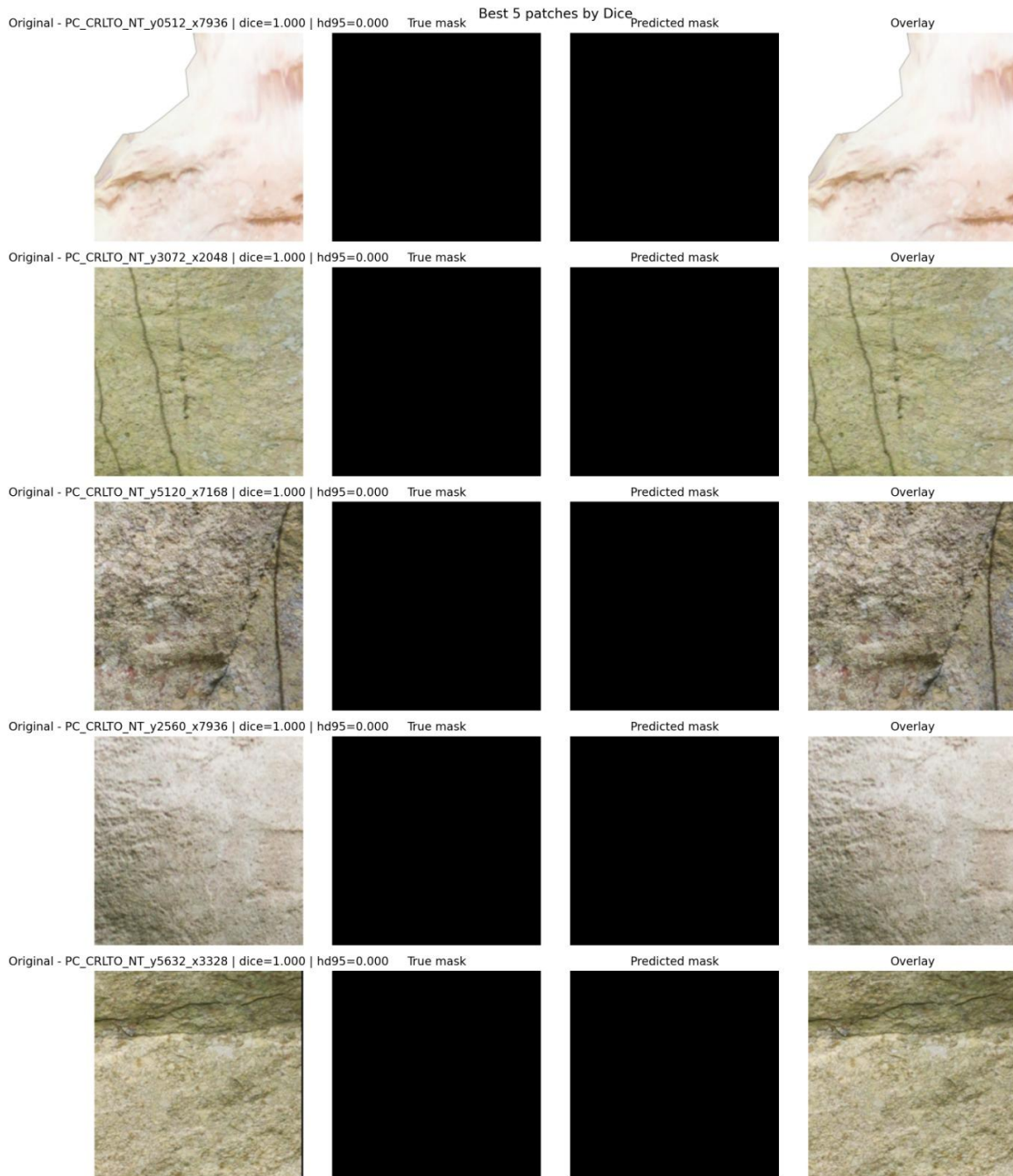


Figura 22. Ejemplos de mejores predicciones del modelo U-Net con undersampling adaptativo según la métrica Dice para la clase plantas superiores.

Por otro lado, los peores casos presentados en la Fig. 18 y la Fig. 19 permiten observar con mayor claridad las limitaciones actuales del modelo. En varios ejemplos se evidencian falsos negativos, donde la máscara real contiene regiones asociadas a plantas superiores, pero el modelo predice únicamente fondo. También se observan falsos positivos y segmentaciones poco precisas en algunos parches, especialmente cuando existen texturas, sombras o variaciones de color que pueden confundirse con la clase objetivo.

Estos errores indican que, aunque el undersampling obliga al modelo a prestar mayor atención a la clase minoritaria, todavía no logra aprender fronteras suficientemente precisas ni diferenciar de manera robusta entre vegetación real y patrones visuales similares presentes en la superficie rocosa.

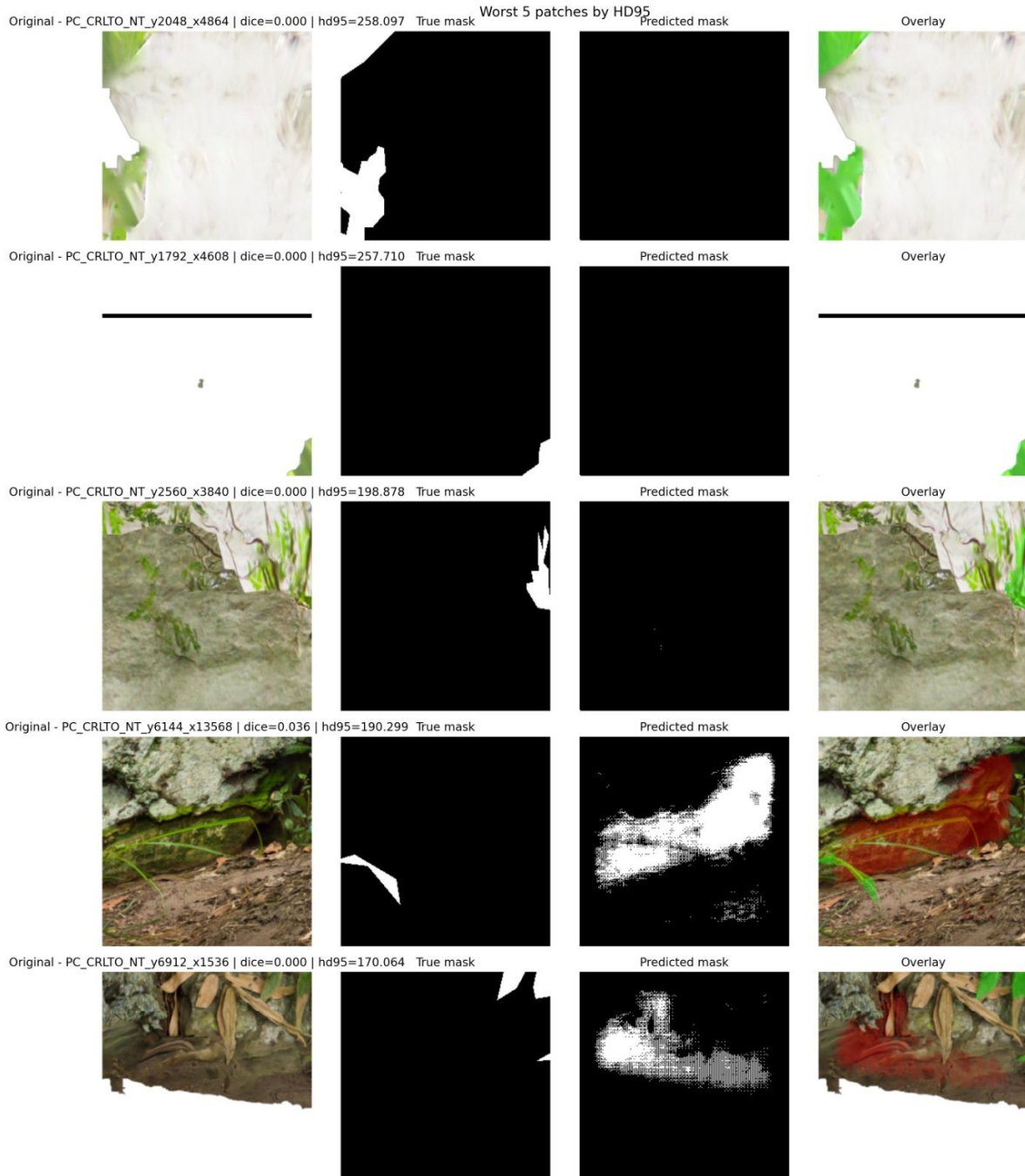


Figura 23. Ejemplos de peores predicciones del modelo U-Net con undersampling adaptativo según la métrica HD95 para la clase plantas superiores.

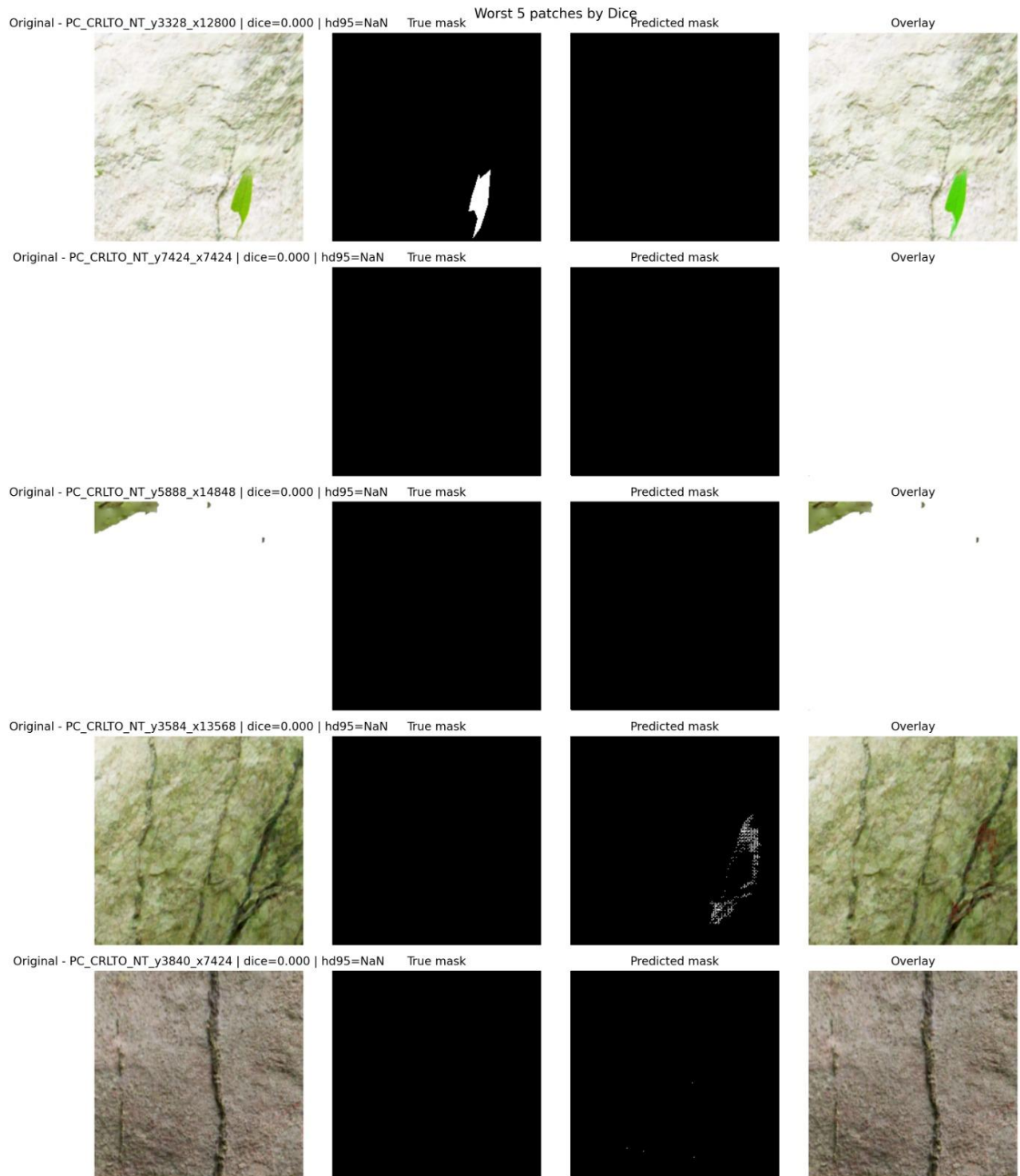


Figura 24. Ejemplos de peores predicciones del modelo U-Net con undersampling adaptativo según la métrica Dice para la clase plantas superiores.

La métrica HD95 continuó siendo especialmente relevante en esta segunda etapa, debido a que permitió identificar errores espaciales que no siempre son evidentes al observar únicamente Dice o IoU. En algunos casos, aunque el modelo detecta parcialmente una región asociada a deterioro, la predicción se encuentra desplazada o presenta una forma considerablemente diferente respecto a la máscara real. Esto genera valores elevados de HD95 y confirma que la evaluación de modelos de segmentación no debe limitarse al solapamiento entre máscaras, sino que también debe considerar la localización y precisión de los bordes. Este aspecto es particularmente importante en el contexto del arte rupestre, donde pequeñas diferencias espaciales pueden afectar la interpretación del deterioro y su utilidad para procesos de conservación.

En conjunto, los resultados de esta segunda etapa muestran que la estrategia de undersampling adaptativo cumple parcialmente su propósito principal: reducir el sesgo del modelo hacia la clase mayoritaria y forzarlo a aprender representaciones asociadas a la clase plantas_superiores. Sin embargo, también pone en evidencia nuevas limitaciones relacionadas con la precisión de las predicciones, el aumento de falsos positivos y la dificultad para delimitar correctamente regiones pequeñas o visualmente ambiguas. Por tanto, esta fase experimental no debe interpretarse como un resultado final, sino como un paso intermedio dentro del proceso de ajuste del modelo. A partir de estos hallazgos, se considera necesario continuar con estrategias complementarias, como el uso de funciones de pérdida más adecuadas para datos desbalanceados, ajustes en el umbral de decisión, técnicas de regularización y evaluaciones separadas sobre parches positivos, con el fin de obtener una medición más robusta y una segmentación más precisa de los bioindicadores de deterioro.

DISCUSIÓN

Los resultados preliminares evidencian que la construcción manual de máscaras representa una de las etapas más complejas y demandantes dentro de proyectos de segmentación semántica aplicados al patrimonio cultural. A diferencia de los conjuntos de datos convencionales utilizados en visión por computador, el presente trabajo requirió un alto grado de intervención manual debido a las características particulares de los mapas de deterioro suministrados por el ICANH. Estos mapas no correspondían directamente a máscaras binarias listas para entrenamiento, sino a representaciones gráficas con simbología especializada, lo que hizo necesario desarrollar un proceso manual de interpretación, delimitación y conversión de las regiones de deterioro. Esta situación incrementó considerablemente los tiempos de preparación del dataset y obligó a tomar decisiones metodológicas cuidadosas para garantizar la coherencia entre las imágenes originales, las máscaras generadas y los patches utilizados durante el entrenamiento.

Por otro lado, las limitaciones de hardware influyeron directamente en la velocidad de experimentación y validación de arquitecturas profundas. El uso inicial de recursos computacionales limitados dificultó la ejecución de entrenamientos extensos y redujo la posibilidad de comparar rápidamente diferentes configuraciones del modelo. Sin embargo, estas dificultades también permitieron fortalecer el componente metodológico del proyecto, ya que hicieron necesario optimizar procesos, organizar el código de manera modular, reducir cargas computacionales innecesarias y evaluar alternativas viables para continuar el entrenamiento. En este sentido, las restricciones técnicas no solo representaron una limitación operativa, sino también una oportunidad para mejorar la planeación experimental y la eficiencia general del flujo de trabajo.

El acompañamiento del director del proyecto fue fundamental para complementar el desarrollo teórico y metodológico, especialmente en aspectos relacionados con la validación experimental, la selección de métricas y el análisis de generalización del modelo. A partir de estas orientaciones fue posible identificar riesgos metodológicos relevantes, como el data leakage derivado de la similitud espacial entre patches de una misma imagen, así como la necesidad de utilizar métricas complementarias que permitieran evaluar no solo el solapamiento entre máscaras, sino también la precisión espacial de las predicciones. Este acompañamiento permitió fortalecer la interpretación de los resultados y evitar conclusiones apresuradas basadas únicamente en métricas globales.

Los resultados preliminares del entrenamiento con U-Net permitieron evidenciar que el desbalance entre parches positivos y negativos influye de manera importante en la interpretación de las métricas globales. Aunque Dice e IoU mostraron valores altos en numerosos casos, el análisis visual de las predicciones y la revisión de la métrica HD95 permitieron identificar errores relevantes en la segmentación de regiones con deterioro real. Esta situación confirma que, en problemas de segmentación aplicados a patrimonio cultural, no basta con reportar métricas globales, ya que estas pueden verse sobreestimadas cuando existe una alta proporción de patches sin presencia de la clase objetivo. Por esta razón, resulta necesario analizar separadamente los casos positivos, los falsos positivos, los falsos negativos y la precisión espacial de las predicciones, especialmente cuando el deterioro ocupa regiones pequeñas dentro de la imagen. La segunda etapa experimental, basada en undersampling adaptativo, permitió comprobar que el balanceo del conjunto de entrenamiento modifica de manera importante el comportamiento del modelo. Aunque las métricas globales disminuyeron frente al escenario sin balanceo, esta reducción no debe interpretarse

únicamente como una pérdida de rendimiento, sino como una evaluación más realista del problema. Al reducir la cantidad de parches negativos durante el entrenamiento, el modelo comenzó a prestar mayor atención a la clase minoritaria; sin embargo, esto también generó nuevos retos asociados a falsos positivos, errores de localización y segmentaciones poco precisas. Este resultado evidencia que el balanceo de datos es necesario, pero no suficiente, ya que el modelo requiere estrategias adicionales para mejorar la delimitación de las regiones afectadas y reducir la detección errónea de texturas similares al deterioro.

En conjunto, los resultados obtenidos muestran que el proyecto se encuentra en una etapa experimental coherente con la complejidad del problema abordado. La implementación inicial de U-Net permitió establecer una línea base funcional, mientras que el análisis del desbalance y la aplicación de undersampling adaptativo permitieron comprender mejor las limitaciones reales del modelo. Estos hallazgos sugieren que las siguientes fases deben orientarse hacia el ajuste de funciones de pérdida más adecuadas para clases desbalanceadas, la evaluación diferenciada sobre patches positivos, el ajuste del umbral de decisión y la exploración de estrategias de regularización que permitan mejorar la precisión espacial de las predicciones. De esta manera, la discusión de los resultados no solo evidencia las limitaciones actuales, sino que también justifica de forma clara las decisiones metodológicas necesarias para continuar fortaleciendo el modelo.

8.1 Problemas computacionales y limitaciones técnicas

Uno de los aspectos más complejos enfrentados durante el desarrollo del proyecto estuvo relacionado con las restricciones de infraestructura computacional disponibles para el entrenamiento de modelos de aprendizaje profundo. Inicialmente, los entrenamientos fueron ejecutados utilizando CPU en entornos locales y virtualizados, lo que generó tiempos de procesamiento extremadamente largos. En algunos casos, una sola corrida experimental podía tardar varios días incluso utilizando un número reducido de épocas. Debido a esta situación, se inició un proceso de exploración de alternativas con aceleración GPU. Entre las pruebas realizadas se incluyen:

- Uso de una máquina virtual institucional equipada con GPU NVIDIA L40-4Q.
- Validaciones de acceso CUDA mediante TensorFlow y PyTorch.
- Configuraciones utilizando entornos Citrix.
- Evaluaciones preliminares con WSL2.
- Exploración de servidores Linux externos.

No obstante, durante las pruebas se identificó que el entorno Citrix no exponía adecuadamente la GPU para tareas de cómputo intensivo. Asimismo, se encontraron incompatibilidades relacionadas con virtualización y soporte CUDA en Windows. Estas limitaciones obligaron a replantear parcialmente la estrategia experimental, considerando alternativas como Google Colab. , Kaggle Notebooks, Servidores Linux externos, Solicitudes institucionales de préstamo de recursos computacionales.

Aunque estas dificultades impactaron el cronograma inicial, también contribuyeron a fortalecer el diseño metodológico del proyecto, obligando a optimizar procesos, modularizar el código y mejorar la eficiencia general del flujo de trabajo. De igual manera, el acompañamiento del director del proyecto permitió complementar el componente teórico y experimental, especialmente en aspectos relacionados con validación metodológica, control de data leakage y definición de métricas robustas para evaluación del modelo. Uno de los principales desafíos del proyecto estuvo relacionado con las limitaciones computacionales disponibles para el entrenamiento de modelos de aprendizaje profundo, las pruebas iniciales ejecutadas en CPU generaron tiempos de entrenamiento excesivamente largos, alcanzando incluso varios días para configuraciones básicas y pocas épocas. Esta situación impactó directamente la velocidad de experimentación y comparación entre arquitecturas.

Con el objetivo de solucionar este inconveniente, se realizaron múltiples intentos de acceso a infraestructura con GPU, incluyendo una máquina virtual de la universidad equipada con una tarjeta NVIDIA L40-4Q. Sin embargo, durante las pruebas realizadas mediante Citrix se identificó que el entorno virtualizado no exponía la GPU para tareas de cómputo CUDA. Adicionalmente, se presentaron incompatibilidades entre TensorFlow y el entorno Windows utilizado, particularmente relacionadas con soporte GPU en versiones recientes sin WSL2 habilitado. También se encontraron restricciones de virtualización y permisos asociados al entorno institucional.

Debido a estas limitaciones, fue necesario explorar alternativas externas como Google Colab, Kaggle Notebooks y posibles servidores Linux con soporte CUDA habilitado. Paralelamente, se adelantaron gestiones para solicitar apoyo institucional y préstamo de recursos computacionales que permitieran continuar el entrenamiento de manera más eficiente. A pesar de estas dificultades, el proyecto logró avanzar de manera constante gracias a la optimización del pipeline, la reorganización metodológica y el acompañamiento técnico del director del trabajo de grado.

CONCLUSIONES Y TRABAJOS FUTUROS

9.1 Conclusiones

El desarrollo del presente proyecto permitió avanzar de manera significativa en la construcción de una metodología orientada a la detección automática de bioindicadores de deterioro presentes en imágenes de arte rupestre mediante técnicas de aprendizaje profundo y visión por computador. A lo largo del trabajo se logró consolidar un flujo metodológico funcional para el procesamiento de imágenes de alta resolución, la construcción manual de máscaras binarias y la preparación de conjuntos de datos supervisados destinados al entrenamiento de modelos de segmentación semántica. Asimismo, el proyecto permitió estructurar una base experimental organizada y reproducible, integrando procesos de generación de patches, validación visual mediante overlays, organización de metadatos y automatización parcial de tareas de carga, entrenamiento y evaluación utilizando Python, TensorFlow y Keras.

Uno de los principales aportes del trabajo radica en la integración de herramientas de inteligencia artificial con problemáticas asociadas a la conservación del patrimonio arqueológico colombiano, evidenciando el potencial de la ciencia de datos como mecanismo de apoyo para procesos de monitoreo patrimonial no invasivo. A nivel experimental, la implementación inicial de una arquitectura U-Net permitió establecer una línea base funcional para la segmentación de la clase plantas superiores. Sin embargo, el análisis detallado de las métricas Dice, IoU y HD95, junto con la inspección visual de las predicciones, permitió identificar que los resultados globales podían verse sobreestimados debido al alto desbalance entre parches positivos y negativos. Este hallazgo fue relevante, ya que permitió comprender que un desempeño aparentemente alto no necesariamente implica una adecuada capacidad para detectar regiones reales de deterioro, especialmente cuando una proporción considerable del conjunto de evaluación corresponde a máscaras vacías o parches sin presencia de la clase objetivo.

De igual manera, la segunda etapa experimental basada en undersampling adaptativo permitió comprobar que el balanceo del conjunto de entrenamiento modifica de forma importante el comportamiento del modelo. Aunque las métricas globales disminuyeron frente al escenario sin balanceo, esta reducción permitió obtener una evaluación más realista del problema, al disminuir el sesgo hacia la clase mayoritaria y obligar al modelo a prestar mayor atención a la clase minoritaria. No obstante, también se evidenciaron nuevas limitaciones relacionadas con falsos positivos, errores de localización y segmentaciones poco precisas, lo cual confirma que el balanceo de datos es necesario, pero no suficiente para resolver completamente el problema. En este sentido, el proyecto permitió establecer una base técnica y metodológica sólida para futuras etapas experimentales, particularmente en aspectos relacionados con control de data leakage, selección de métricas robustas, evaluación diferenciada de parches positivos y negativos, y ajuste de estrategias de entrenamiento para clases desbalanceadas.

A pesar de las limitaciones computacionales enfrentadas durante el desarrollo, especialmente aquellas relacionadas con acceso restringido a recursos GPU, tiempos elevados de entrenamiento y dificultades de compatibilidad entre entornos virtualizados y herramientas de aceleración CUDA, el proyecto logró mantener un avance constante y metodológicamente coherente con los objetivos planteados. Estas dificultades no solo permitieron fortalecer el diseño experimental y la organización del trabajo, sino que también contribuyeron al desarrollo de competencias investigativas relacionadas con procesamiento digital de imágenes, aprendizaje profundo, validación experimental y estructuración de proyectos de investigación aplicada. En este sentido, el trabajo realizado constituye una base relevante para futuras investigaciones orientadas al uso de inteligencia artificial en contextos de conservación patrimonial y análisis automatizado de deterioros arqueológicos.

9.2 Trabajos Futuros

Como continuación del trabajo desarrollado, se plantea inicialmente completar las fases experimentales restantes asociadas al entrenamiento y comparación de múltiples arquitecturas de segmentación semántica. Aunque durante esta etapa se priorizó la implementación inicial de U-Net, el entorno metodológico construido permite incorporar modelos más avanzados como ResUNet, Attention U-Net, DeepLabv3+ y arquitecturas híbridas basadas en Transformers, con el propósito de evaluar posibles mejoras en capacidad de generalización y precisión de segmentación. Del mismo modo, se propone fortalecer las estrategias de aumento de datos (data augmentation), buscando mejorar la variabilidad del conjunto de entrenamiento y reducir problemas asociados al número limitado de paneles actualmente disponibles.

Otro aspecto importante para trabajos futuros corresponde a la expansión progresiva del dataset, incorporando nuevos paneles rupestres y ampliando la cantidad de categorías de deterioro disponibles para entrenamiento supervisado. Actualmente, una de las principales limitaciones metodológicas está relacionada con la distribución desigual de deterioros entre los paneles existentes, situación que dificulta la implementación de estrategias estrictas de validación inter-panel. Por esta razón, la inclusión de nuevas imágenes anotadas permitiría construir modelos multicategoría más robustos, mejorar la capacidad de generalización y evaluar de manera más rigurosa el desempeño del sistema frente a escenarios reales de conservación patrimonial.

Finalmente, se propone explorar posibles aplicaciones prácticas derivadas del sistema desarrollado, incluyendo plataformas web de monitoreo, herramientas de apoyo para especialistas en conservación y sistemas de análisis temporal orientados al seguimiento de la evolución del deterioro en diferentes periodos. Asimismo, sería relevante validar experimentalmente el modelo junto a expertos en conservación patrimonial y arqueología, permitiendo contrastar los resultados automáticos obtenidos con evaluaciones realizadas por especialistas. Este tipo de integración interdisciplinaria podría contribuir significativamente al desarrollo de herramientas computacionales aplicadas al patrimonio cultural colombiano y abrir nuevas líneas de investigación relacionadas con inteligencia artificial, conservación digital y documentación arqueológica automatizada.

Bibliografía

- [1] G. Muñoz Castiblanco, «Estética amazónica y discusiones contemporáneas: El arte rupestre de la serranía de La Lindosa, Guaviare - Colombia», *Calle 14 Rev. Investig. En El Campo Arte*, vol. 15, n.º 27, pp. 14-39, oct. 2019, doi: 10.14483/21450706.15406.
- [2] J. V. Becerra, *Pinturas rupestres de la vereda Nuevo Tolima, San José del Guaviare*. Bogotá, Colombia: OIM, Misión Colombia, 2019.
- [3] J. L. Yela y M. Sameño Puerto, «Los insectos y el biodeterioro del Patrimonio Histórico Cultural», *Rev. PH*, p. 67, mar. 1997, doi: 10.33349/1997.18.460.
- [4] G. Caneva, M. P. Nugari, y O. Salvadori, *Biology in the conservation of works of art*. Rom: International Centre for the study of the Preservation, 1991.
- [5] N. Kheradmandi y V. Mehranfar, «A critical review and comparative study on image segmentation-based techniques for pavement crack detection», *Constr. Build. Mater.*, vol. 321, p. 126162, feb. 2022, doi: 10.1016/j.conbuildmat.2021.126162.
- [6] O. Ronneberger, P. Fischer, y T. Brox, «U-Net: Convolutional Networks for Biomedical Image Segmentation», en *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, y A. F. Frangi, Eds., en *Lecture Notes in Computer Science*, vol. 9351., Cham: Springer International Publishing, 2015, pp. 234-241. doi: 10.1007/978-3-319-24574-4_28.
- [7] ICANH, Escuela Nacional de Conservación, Restauración y Museografía Manuel Del Castillo Negrete, y INAH, «Preserving the past: preventive conservation on World Heritage Site Chiribiquete National Park and its buffer zone (Colombia) Diagnostico del estado de conservación de sitios rupestres en el Area Nacional Protegida Arqueológica». 2023.
- [8] UNESCO, «Recomendación de la UNESCO sobre la Ciencia Abierta», UNESCO, 2021. doi: 10.54677/YDOG4702.
- [9] A. Zerboni *et al.*, «The Sustainability of Rock Art: Preservation and Research», *Sustainability*, vol. 14, n.º 10, p. 6305, may 2022, doi: 10.3390/su14106305.
- [10] G. Caneva, M. P. Nugari, y O. Salvadori, *Biology in the conservation of works of art*. Rom: International Centre for the study of the Preservation, 1991.
- [11] V. Bayarri, M. A. Sebastián, y S. Ripoll, «Hyperspectral Imaging Techniques for the Study, Conservation and Management of Rock Art», *Appl. Sci.*, vol. 9, n.º 23, p. 5011, nov. 2019, doi: 10.3390/app9235011.
- [12] T. Hastie, R. Tibshirani, y J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. en Springer series in statistics. New York, NY: Springer, 2009.
- [13] I. Goodfellow, A. Courville, y Y. Bengio, *Deep learning*. en Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2016.
- [14] J. Long, E. Shelhamer, y T. Darrell, «Fully Convolutional Networks for Semantic Segmentation», 2014, *arXiv*. doi: 10.48550/ARXIV.1411.4038.
- [15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, y H. Adam, «Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation», 2018, *arXiv*. doi: 10.48550/ARXIV.1802.02611.
- [16] K. He, G. Gkioxari, P. Dollar, y R. Girshick, «Mask R-CNN», en *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, oct. 2017, pp. 2980-2988. doi: 10.1109/ICCV.2017.322.
- [17] N. Tasselli y F. Maietti, «BIM FOR EXISTING HERITAGE: FROM 3D INTEGRATED SURVEY TO PARAMETRIC MODELING FOR REFURBISHMENT AND MANAGEMENT», *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLVI-2/W1-2022, pp. 481-488, feb. 2022, doi: 10.5194/isprs-archives-XLVI-2-W1-2022-481-2022.
- [18] S. Kaufman, S. Rosset, C. Perlich, y O. Stitelman, «Leakage in data mining: Formulation, detection, and avoidance», *ACM Trans. Knowl. Discov. Data*, vol. 6, n.º 4, pp. 1-21, dic. 2012, doi:

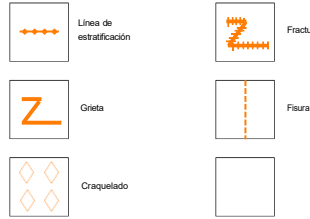
10.1145/2382577.2382579.

- [19] R. Szeliski, *Computer Vision: Algorithms and Applications*. en Texts in Computer Science. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-030-34372-9.
- [20] S. Varma y R. Simon, «Bias in error estimation when using cross-validation for model selection», *BMC Bioinformatics*, vol. 7, n.º 1, p. 91, dic. 2006, doi: 10.1186/1471-2105-7-91.
- [21] DIEGO ALEJANDRO RODRÍGUEZ FERNÁNDEZ, «Análisis de la integración SIG, topografía y digitalización vectorial de imágenes para la gestión sostenible de la conservación y documentación de arte rupestre en el área arqueológica protegida Serranía de la Lindosa», Universidad Distrital Francisco Jose De Caldas, Bogota, 2024. [En línea]. Disponible en: <https://repository.udistrital.edu.co/items/533086d4-074d-4f41-a57b-8b5bb9742bb7>
- [22] M. Ricca *et al.*, «Multidisciplinary Approach to Characterize Archaeological Materials and Status of Conservation of the Roman Thermae of Reggio Calabria Site (Calabria, South Italy)», *Appl. Sci.*, vol. 10, n.º 15, p. 5106, jul. 2020, doi: 10.3390/app10155106.

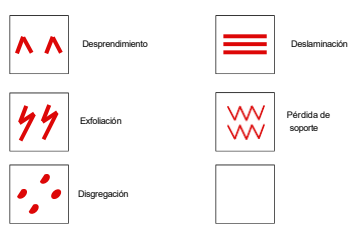
Anexos.

SOPORTE PÉTREO

GRIETAS Y DEFORMACIONES [Color naranja]



DESPRENDIMIENTOS [Color rojo]

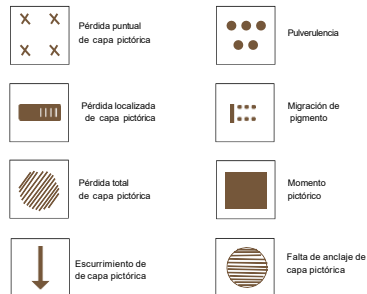


SOPORTE PÉTREO

EFFECTOS POR PÉRDIDA DE MATERIAL [Color morado]

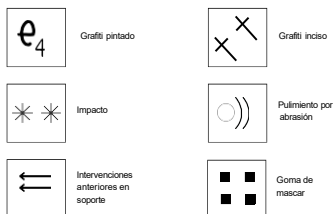


CAPA PICTÓRICA [Color café]



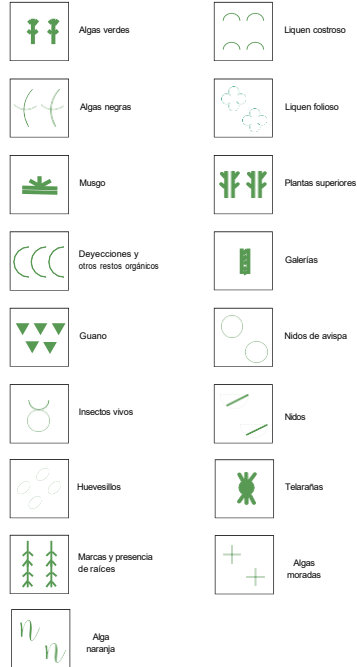
GENERALES

FENÓMENOS ANTROPOGÉNICOS [Color negro]



GENERALES

FENÓMENOS BIOLÓGICOS [Color verde]



GENERALES

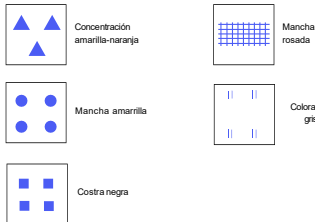
FENÓMENOS SALINOS [Color azul]



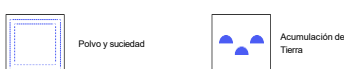
FENÓMENOS HÍDRICOS [Color azul]

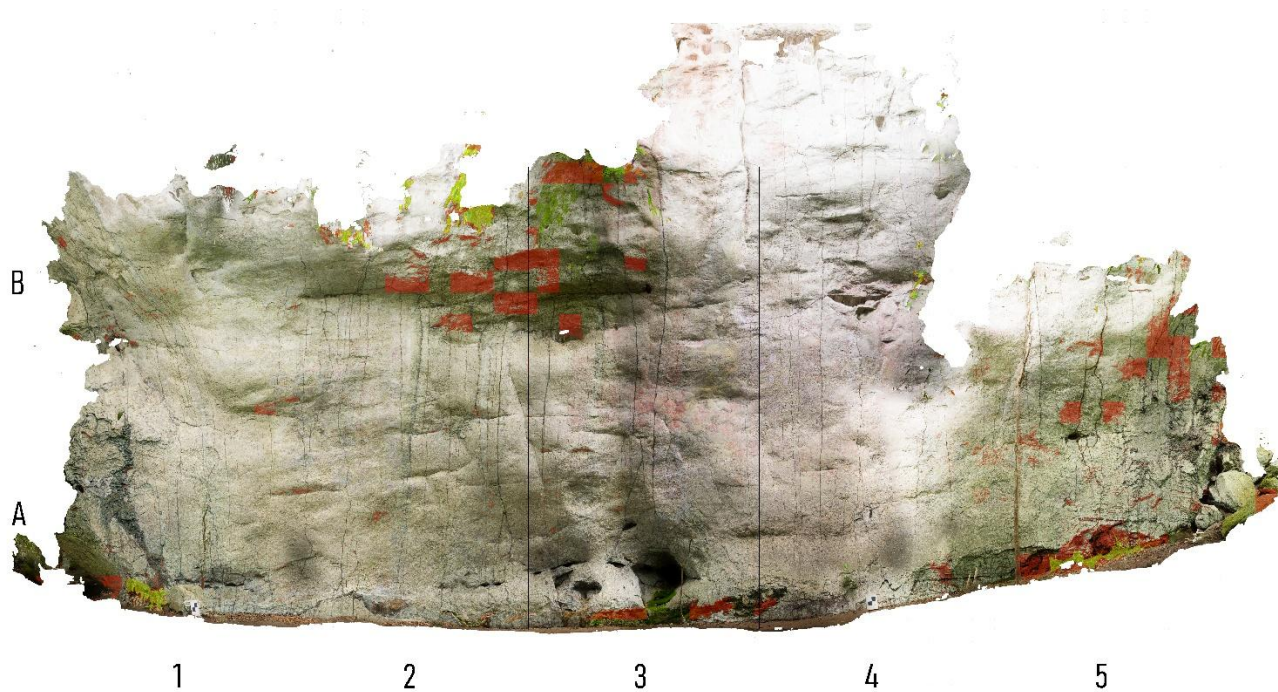


FENÓMENOS CROMÁTICOS [Color azul]

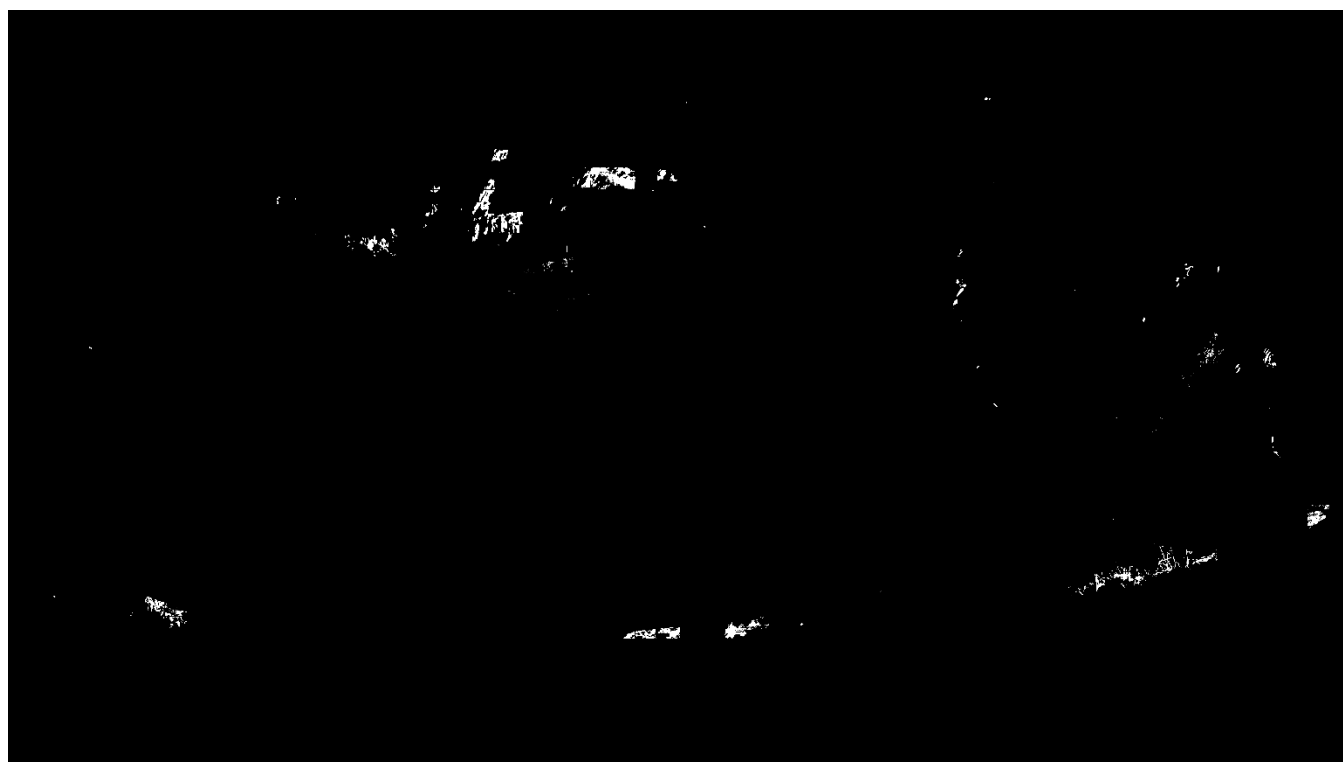


FENÓMENOS DEPOSICIONALES [Color azul]





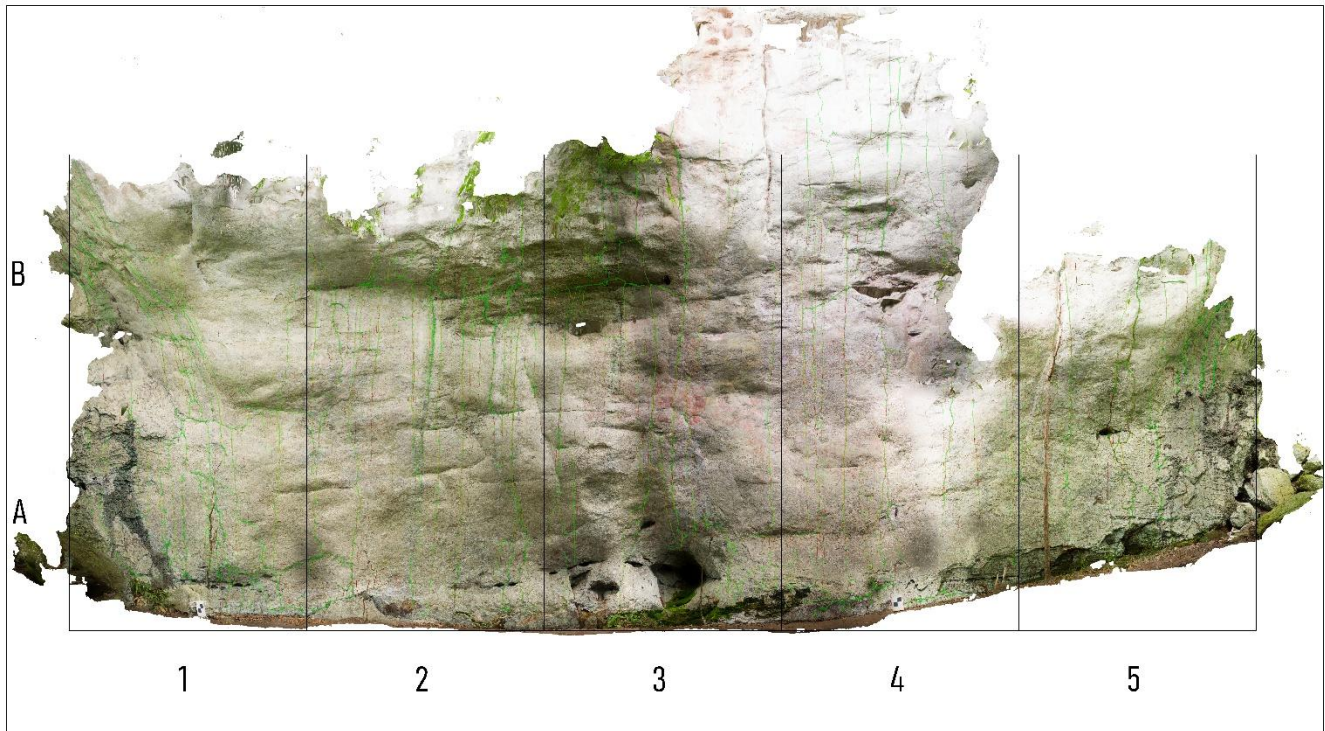
Anexo 2. Superposición en imagen original de mascara predicha (Rojo) vs. Mascara manual (Verde) en deterioro plantas superiores.



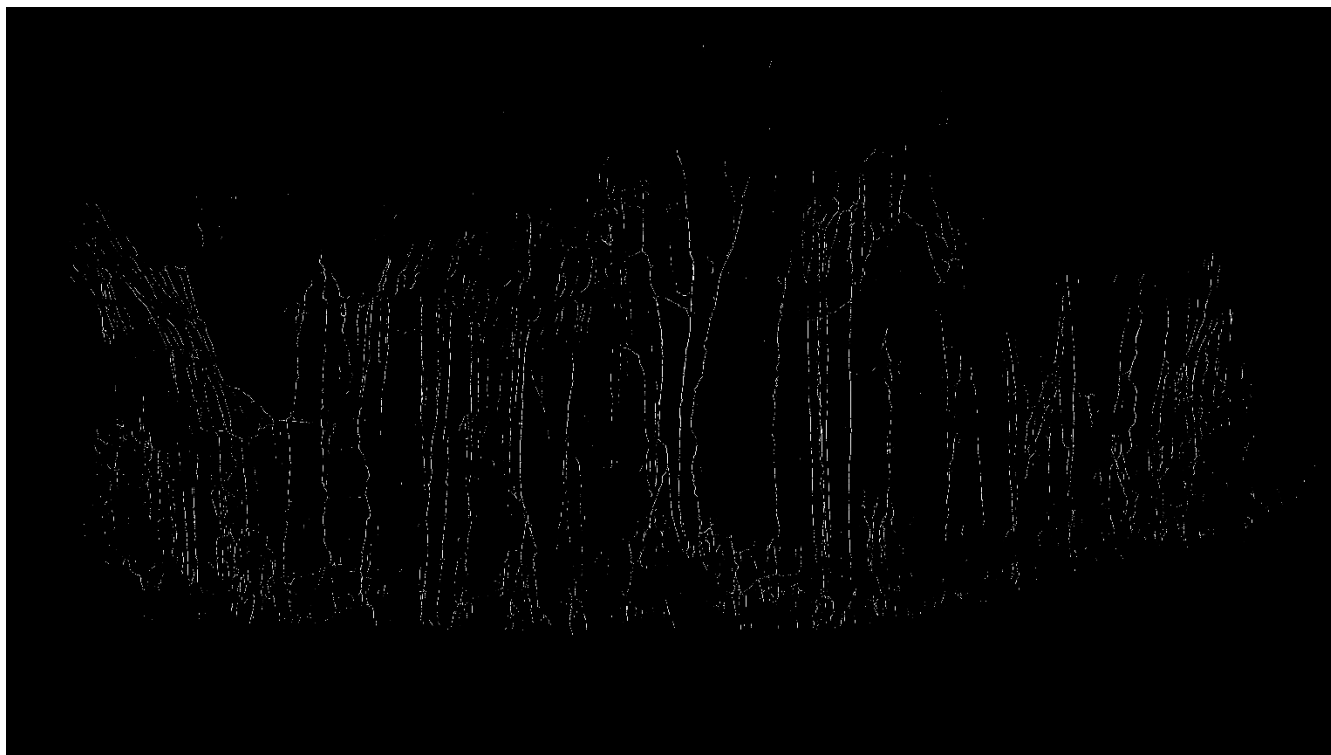
Anexo 3. Mascara predicha para deterioro plantas superiores.



Anexo 4. Mascara manual para deterioro plantas superiores.



Anexo 5. Superposición en imagen original de mascara predicha (Rojo) vs. Mascara manual (Verde) en deterioro galerías.



Anexo 6. Mascara predicha para deterioro galerías.



Anexo 7. Mascara manual para deterioro galerías.