

SISTEMA EXPERTO BASADO EN VISIÓN ARTIFICIAL PARA LA CLASIFICACIÓN AUTOMÁTICA DE RESIDUOS DE HORMIGÓN A PARTIR DE LA IDENTIFICACIÓN DE CARBONATACIÓN EMPLEANDO INDICADOR DE PH INOCUO

Rafael Hidalgo-Gato García¹, Mario González Martínez¹, Ruth García García¹, Alejandro López García¹
Daniela Eugenia Angulo Ramírez², Antonio Jesús Martínez-Espinosa², Francisco Javier Benito Saorín³

¹ Centro Tecnológico CTC, Santander, España, rhidalgo@centrotecnologicocctc.com

² Construcciones Urdecon S.A., Murcia, España, amartinez@urdecon.es

³ Universidad de Alicante, Alicante, España, franciscojavier.benito@ua.es

Resumen

La carbonatación del hormigón constituye uno de los principales procesos de degradación de los materiales cementicios, al afectar tanto a la durabilidad del material como a la protección de las armaduras embebidas. Este fenómeno implica la reacción del dióxido de carbono atmosférico con los hidróxidos alcalinos presentes en la matriz cementicia, en particular con el hidróxido cálcico, lo que da lugar a la formación de carbonato cálcico y a una disminución progresiva del pH. La detección precisa de este fenómeno es fundamental tanto para la evaluación de la durabilidad de estructuras como para el análisis del potencial de reutilización de elementos de hormigón reciclado. Se presenta un sistema experto de visión artificial para la detección automática y segmentación de zonas carbonatadas y no carbonatadas en fragmentos de hormigón destinados al reciclaje, empleando como referencia visual un indicador de pH inocuo formulado con cúrcuma. El sistema integra una cámara multiespectral RGB/NIR con iluminación activa controlada, un modelo de segmentación de la familia YOLO entrenado con un conjunto de imágenes propias anotadas en 2 clases operativas vinculadas al estado de carbonatación y un criterio de fusión semántico-geométrica permitiendo generar un registro individual con etiqueta, porcentaje dominante y estado de validez por cada fragmento analizado. La trazabilidad experimental se gestiona mediante MLflow, que registra hiperparámetros, métricas y artefactos de cada iteración de entrenamiento. Los resultados muestran que el enfoque proporciona una clasificación objetiva, trazable y reproducible del estado de carbonatación por fragmento, con un comportamiento conservador ante señales cromáticas desconocidas y con métricas operativas estables bajo condiciones de laboratorio controladas. La metodología establece las bases metodológicas para la integración de este tipo de módulos en flujos automatizados de valorización de residuos de hormigón.

Palabras Clave: visión artificial, YOLO, residuos, carbonatación.

1 Introducción

La gestión de residuos de construcción y demolición (RCD) representa uno de los retos más relevantes para la economía circular en el sector de la construcción. El hormigón constituye la fracción predominante de estos residuos y su reutilización como árido reciclado exige criterios objetivos que permitan separar el material según su estado químico superficial [1], [2]. La carbonatación, resultado de la reacción del dióxido de carbono atmosférico con los hidróxidos alcalinos de la matriz cementicia, reduce el pH del material desde valores próximos a 12-13 en zonas no carbonatadas hasta rangos cercanos a 8-9 en zonas completamente carbonatadas [1].

Esta diferencia de alcalinidad condiciona la reutilización del árido reciclado, en particular cuando las especificaciones técnicas establecen límites sobre la presencia de carbonatación superficial o definen criterios de compatibilidad con nuevas matrices cementadas [2], [3]. La determinación visual del frente de carbonatación se ha realizado tradicionalmente mediante indicadores de pH aplicados sobre superficies recién expuestas por fractura. La fenolftaleína ha sido el reactivo de referencia por su cambio cromático nítido entre pH 9 y 10.5, pero su clasificación como sustancia potencialmente cancerígena (categoría 2) ha impulsado la búsqueda de alternativas con menor riesgo operativo [4], [5]. Indicadores de base natural, como los formulados a partir de cúrcuma o antocianinas, presentan una transición cromática visible en el rango de pH relevante para la detección de carbonatación y ofrecen ventajas en cuanto a inocuidad, biodegradabilidad y compatibilidad con protocolos de ensayo de alta frecuencia [4], [5], [6]. La automatización del análisis colorimétrico asociado al indicador resulta especialmente necesaria cuando los residuos no se presentan como probetas regulares, sino como conjuntos de fragmentos con geometría irregular, árido expuesto, porosidad abierta y texturas superficiales heterogéneas. En estas condiciones, la lectura manual pierde repetibilidad y los métodos basados en un único umbral de color son sensibles a variaciones de iluminación, estado superficial y operador [7], [8]. Esta situación justifica abordar el problema como una tarea integrada de adquisición controlada, segmentación semántica y decisión por elemento físico. El objetivo del presente trabajo es desarrollar y validar un sistema 2D multispectral para la detección y segmentación automática de zonas carbonatadas y no carbonatadas en fragmentos de hormigón, empleando un indicador de pH inocuo basado en cúrcuma como referencia cromática y un modelo de segmentación de la familia YOLO como núcleo de inferencia. La aportación del trabajo se articula en 4 componentes diferenciados: un protocolo de aplicación del indicador inocuo caracterizado experimentalmente en términos de composición, método de aplicación y ventana temporal de validez; un módulo de captura RGB/NIR con iluminación activa y geometría de adquisición fija; un modelo YOLO de segmentación (sobre RGB) entrenado sobre clases operativas vinculadas al estado de carbonatación, optimizado mediante un plan sistemático de 105 experimentos gestionados en MLflow; y un criterio de fusión semántico-geométrica que transforma las máscaras de píxel en indicadores cuantitativos trazables por fragmento físico. La Figura 1 resume la arquitectura general del sistema y la relación entre sus componentes principales.

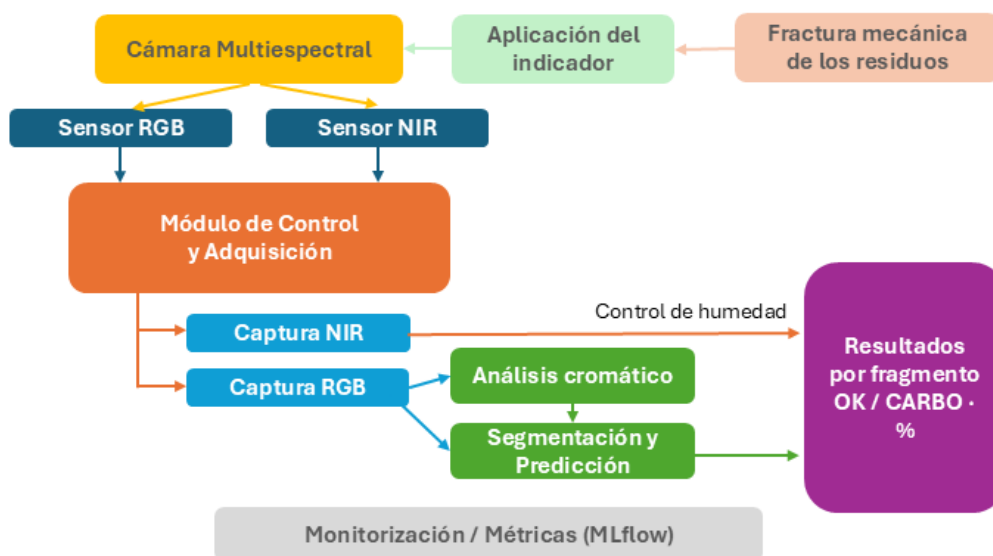


Figura 1. Flujo operativo del sistema experto de visión artificial propuesto.

2 Antecedentes

La medición automatizada de la carbonatación mediante análisis de imagen ha avanzado desde la inspección visual hacia métodos reproducibles basados en segmentación digital. Los trabajos de Choi et al. [7] y Giuliatti et al. [8] demostraron que el análisis de imágenes obtenidas sobre superficies tratadas con fenolftaleína puede reducir la variabilidad del operador y generar medidas consistentes del frente de carbonatación. De forma complementaria, Silveira e Isaia [9] propusieron un procedimiento digital para la delimitación del frente carbonatado y Zhang et al. y Wang et al. revisaron los principales retos asociados a la medida precisa de la profundidad de carbonatación en materiales cementicios [10], [11]. Estos trabajos confirman que el problema no se reduce a detectar un color, sino a resolver una frontera difusa sobre superficies rugosas con gradientes cromáticos, poros, áridos visibles y variaciones locales de humedad. El uso de indicadores alternativos a la fenolftaleína ha sido abordado por Vogler et al. [4], quienes evaluaron varios reactivos inocuos sobre muestras cementicias y por autores que propusieron soluciones basadas en antocianinas [6]. Estos indicadores son viables para discriminar zonas carbonatadas y no carbonatadas, pero su respuesta colorimétrica es más sensible al tiempo de lectura, la humedad superficial y la composición del soporte que la de la fenolftaleína [4], [5]. Esta mayor sensibilidad requiere un flujo de captura y análisis capaz de compensar la variabilidad temporal y superficial del contraste.

En el ámbito de la visión artificial aplicada a materiales cementicios, las redes convolucionales y las arquitecturas encoder-decoder han mostrado mayor capacidad que los métodos clásicos de umbralización para segmentar superficies heterogéneas con contornos irregulares [12], [13]. La familia YOLO ofrece un equilibrio favorable entre velocidad de inferencia y precisión para tareas de detección y segmentación en tiempo real y su aplicación en inspección de estructuras de hormigón ha sido evaluada en detección de fisuras, corrosión y defectos superficiales [14], [15], [16]. Las revisiones recientes de la evolución de YOLO describen una transición progresiva hacia arquitecturas más eficientes con mejoras en fusión multiescala, funciones de pérdida y variantes de segmentación de instancias [16], [17], [18]. Las prácticas de MLOps y herramientas como MLflow proporcionan trazabilidad sobre experimentos, métricas y artefactos, aspecto fundamental cuando el rendimiento del modelo depende de cambios en el conjunto de datos, en los hiperparámetros o en las condiciones de captura [19], [20]. La adquisición multiespectral RGB/NIR ha mostrado utilidad en evaluación de daño superficial en hormigón y en análisis de materiales heterogéneos, al aportar información espectral adicional sobre textura, reflectancia y humedad que no está disponible en el canal visible [21], [22], [23]. El posicionamiento del trabajo respecto al estado del arte se basa en la combinación de 3 elementos no integrados previamente: la caracterización experimental cuantitativa de un indicador inocuo de cúrcuma en términos de ventana temporal de validez y tasa de degradación cromática; el diseño de un módulo de captura RGB/NIR específicamente orientado a resolver el contraste colorimétrico del indicador sobre fragmentos irregulares de RCD; y la integración de un modelo YOLO de segmentación con un criterio de fusión semántico-geométrica que expresa el resultado de clasificación a nivel de fragmento físico en lugar de a nivel de escena global.

3 Materiales y métodos

3.1. Muestras, indicador y protocolo de aplicación

Se emplearon fragmentos de hormigón de demolición de alta densidad (H01), obtenidos mediante fractura mecánica por impacto a partir de piezas de 10-15 cm, que generaron superficies internas recién expuestas de 4-6 cm. La fractura mecánica por impacto se seleccionó frente al corte con herramienta eléctrica debido a que el calor generado por el disco durante el corte

modifica el pH superficial y altera la respuesta cromática del indicador. Las muestras se conservaron en laboratorio bajo condiciones ambientales controladas ($22 \pm 0.5^\circ\text{C}$ y $40 \pm 2\%$ de humedad relativa) y se procesaron inmediatamente tras la fractura para minimizar la carbonatación superficial incipiente. El indicador de pH se formuló a partir de cúrcuma al 0.5 % m/m en etanol al 70 %, seleccionado por su respuesta cromática en medios alcalinos y por su menor criticidad operativa frente a indicadores convencionales como la fenolftaleína [4], [5]. La transición cromática del pigmento curcuminoide ocurre desde tonalidades amarillas en medios ácidos o neutros ($\text{pH} < 8$) hacia naranja-rojizo en medios fuertemente básicos ($\text{pH} > 9$), lo que permite discriminar visualmente las zonas no carbonatadas de las carbonatadas [5], [6]. La comparación directa con fenolftaleína sobre 20 puntos de referencia por probeta mostró desviaciones inferiores al 20 % en todas las mediciones, confirmando la validez del indicador como sustituto operativo. La aplicación se realizó mediante pulverización fina a una distancia aproximada de 200 mm sobre la superficie recién expuesta, con un ángulo de inclinación de 45° y un patrón homogéneo, evitando acumulaciones localizadas del reactivo que pudieran generar falsos patrones de coloración. Tras la aplicación del indicador se estableció un periodo de estabilización de aproximadamente 5 min, supervisado mediante el canal NIR para verificar la ausencia de humedad residual antes de la captura. Los ensayos de degradación temporal del contraste cromático, realizados mediante captura continua durante 8 horas (480 imágenes a 0,0167 Hz), mostraron para la cúrcuma una tasa de degradación de $-7.8\% h^{-1}$ durante las 2 primeras horas y de $-5.2\% h^{-1}$ a partir de ese punto, con una pérdida total del 63 % respecto al valor inicial. La fenolftaleína presentó una pérdida total del 28 % en el mismo periodo y una tasa media de $3.1\% h^{-1}$, lo que supone aproximadamente el doble de estabilidad colorimétrica. Con un umbral operativo fijado en el 70 % de la saturación inicial del indicador inocuo, la ventana de identificación fiable del frente de carbonatación se situó en aproximadamente 3-4 horas. Los ensayos con intervalos de aplicación de 0, 2, 4 y 6 horas posteriores a la fractura, realizados en 2 series (corte eléctrico y fractura manual), confirmaron que el tiempo máximo efectivo de aplicación del indicador es de 4 horas en condiciones de laboratorio controladas. Transcurrido ese intervalo, la carbonatación superficial reduce el pH por debajo del umbral de viraje, limitando la reacción colorimétrica.

3.2. Sistema de captura RGB/NIR

El módulo óptico se configuró con una cámara multispectral JAI FS-1600D-10GE, equipada con arquitectura prismática y doble sensor CMOS que permite la captura simultánea y co-registrada de imágenes RGB y NIR a través de una única óptica. El objetivo empleado fue un TV Zoom Lens 11.5-69 mm F/1.4, con un filtro polarizado MIDOPT PR032-46 en la rosca frontal para reducir reflejos especulares sobre áridos y superficies húmedas. El área útil de medida se fijó en 210×297 mm sobre fondo negro. Considerando una resolución de adquisición de 1440×1080 píxeles, una distancia de trabajo de 653 mm y una focal de 11.5 mm, se obtuvo una resolución espacial de 0.196 mm/px en el plano de medida. La iluminación en el rango visible se resolvió mediante 2 barras LED lineales (LED EFF-FL25) dispuestas de forma simétrica respecto al área de medida, que proporcionan iluminación homogénea y reducen sombras y reflejos especulares. Para el canal NIR se empleó un anillo de iluminación centrado en 850 nm (EFFI-RING 850 nm), coincidente con uno de los picos de sensibilidad del sensor NIR de la cámara. Durante todas las campañas de captura se mantuvieron constantes la geometría de cámara, la focal, el encuadre, la iluminación, el fondo y los parámetros de adquisición, con el fin de reducir la deriva radiométrica entre sesiones. El canal RGB constituyó la señal analítica principal para la inferencia del modelo. El canal NIR se empleó como apoyo visual de comprobación para detectar contenido de humedad superficial residual, sin participar como entrada al modelo de segmentación. La Figura 2 muestra la configuración óptica del sistema.

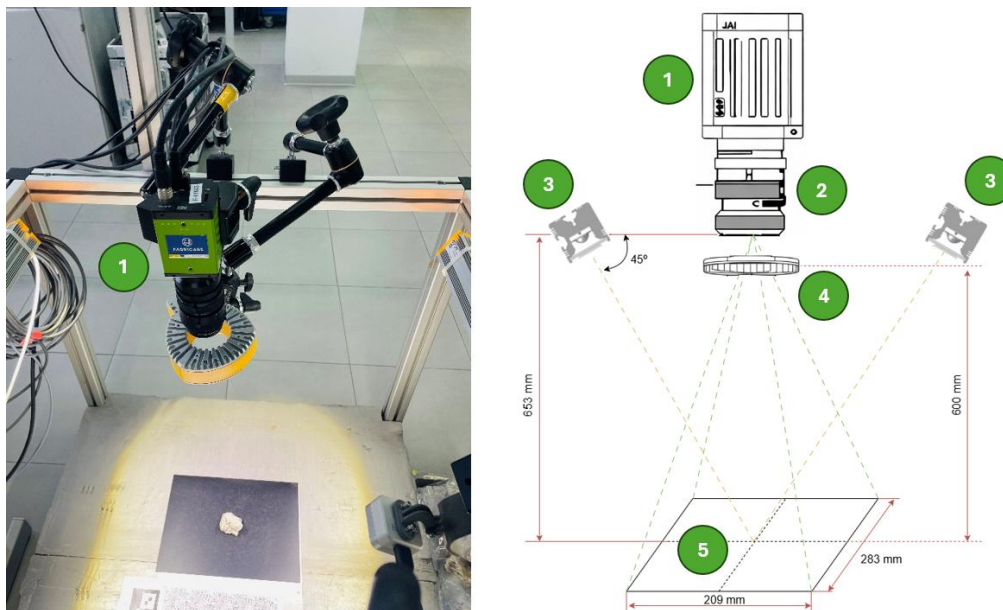


Figura 2. Configuración óptica del sistema de captura multispectral. Cámara multispectral JAI FS-1600D-10GE (RGB VIS y NIR) (1), óptica VF 11.5-69mm (2), Barras de Iluminación LED EFF-FL25 (3), Anillo EFFI-RING 850nm (4) y área de medida de 210 × 297 mm (5).

3.3. Composición del dataset y etiquetado

Durante la fase experimental se generaron 659 imágenes. Tras aplicar criterios de calidad relacionados con iluminación, enfoque, contraste útil y trazabilidad de la muestra, se seleccionó un conjunto inicial de 230 imágenes para el entrenamiento del modelo optimizado (versión v1), distribuidas en 105 imágenes para entrenamiento, 45 para validación, 50 para test y 30 para inferencia cualitativa. El dataset se amplió posteriormente a 308 imágenes (versión v2), incorporando nuevas capturas que diversificaban las condiciones de iluminación, la morfología de los fragmentos y el grado de saturación del indicador, 178 para entrenamiento, 50 para validación, 50 para test y 30 para inferencia. Las clases operativas se definieron como OK, para regiones no carbonatadas con respuesta cromática del indicador y CARBO, para regiones carbonatadas o con ausencia dominante de respuesta alcalina. La elección de una formulación binaria respondió a la baja robustez observada al intentar separar grados cromáticos intermedios en superficies de hormigón fracturado, pequeñas variaciones de iluminación, rugosidad, humedad o espesor del indicador producen transiciones tonales próximas entre sí, por lo que la formulación binaria permite una tarea más estable y alineada con una decisión operativa por fragmento. Las imágenes con bajo contraste, iluminación no homogénea, sombras proyectadas o fragmentos parcialmente fuera del área útil fueron descartadas antes del entrenamiento.

4 Sistema de segmentación e inferencia

4.1. Modelo YOLO y plan de entrenamiento

El núcleo de inferencia se implementó con el modelo YOLO11n-seg, variante nano de la arquitectura YOLOv11 con cabeza de segmentación de instancias, inicializado a partir de los pesos públicos YOLO11n. Esta inicialización aporta representaciones jerárquicas de bajo nivel transferibles al dominio del hormigón sin introducir sesgos de dominio específico. La selección

de YOLO11n-seg sobre otras variantes de las familias YOLOv8 y YOLO11 resultó de un análisis preliminar de 7 iteraciones comparativas, en el que la variante nano con cabeza de segmentación ofreció el mejor equilibrio entre calidad de contornos y velocidad de inferencia sobre el dataset inicial. El modelo se entrenó sobre las clases OK y CARBO mediante un plan experimental sistemático de 105 ejecuciones organizadas en 6 bloques temáticos y gestionadas íntegramente con MLflow, optimización del aprendizaje, regularización y scheduling, tratamiento del color, aumentaciones geométricas, estabilidad y reproducibilidad y barridos de inferencia [20], [24]. El plan se diseñó para evaluar de forma secuencial los grupos de hiperparámetros con mayor impacto sobre la calidad de la segmentación, manteniendo fijos los parámetros no evaluados en cada bloque y utilizando la configuración óptima identificada como punto de partida del bloque siguiente. La Tabla 1 recoge el resultado final de este proceso.

Tabla 1. Bloques del plan experimental de optimización.

Bloque	Parámetros evaluados	Ejecuciones	Resultados
1. Optimización del aprendizaje	lr0, lrf, batch, optimizador, momentum, weight_decay	48	SGD con lr0 = 0.0016 y momentum = 0.94 supera a AdamW en estabilidad de convergencia y calidad geométrica de máscaras
2. Regularización y scheduling	cos_lr, warmup_epochs, patience, epochs	14	Scheduler cosenoidal activo y warmup de 5 épocas reducen la variabilidad en la fase de convergencia inicial
3. Tratamiento del color	hsv_h, hsv_s, hsv_v, bgr	15	Saturación moderada (hsv_s = 0.40) aporta robustez ante variaciones de iluminación sin distorsionar la firma espectral del indicador
4. Aumentaciones geométricas	degrees, fliplr, flipud	13	Rotación conservadora ($\pm 10^\circ$) preserva la coherencia física de la adquisición sin degradar la localización de contornos
5. Estabilidad y reproducibilidad	seed	5	seed = 42 ofrece la mayor consistencia entre ejecuciones con hiperparámetros fijos
6. Barridos de inferencia	conf, loU	10	conf = 0.50 e loU = 0.55 maximizan el equilibrio precisión/sensibilidad en el dominio evaluado
Total		105	

La configuración definitiva del modelo integra el conjunto de hiperparámetros identificado como óptimo a partir del análisis por bloques. La arquitectura YOLO11n-seg se entrenó con una resolución de entrada de 640×640 px, que ofrece el mejor equilibrio entre precisión geométrica y coste computacional, un tamaño de lote de 16 imágenes y un máximo de 200 épocas con parada anticipada controlada por un criterio de patience de 30 épocas sin mejora.

El optimizador seleccionado fue SGD, cuyo análisis comparativo frente a AdamW confirmó su superioridad tanto en estabilidad de las curvas de entrenamiento como en fidelidad espacial de los contornos de las máscaras generadas en inferencia. La tasa de aprendizaje inicial se fijó en lr0 = 0.0016, con un factor de decaimiento lrf = 0.006, momentum = 0.94 y regularización L2 weight_decay = 0.001.

La dinámica de aprendizaje se estabilizó mediante un scheduler cosenoidal (\cos_lr activo) con 5 épocas de calentamiento previo ($warmup_epochs$) y la reproducibilidad entre ejecuciones se garantizó fijando la semilla aleatoria en $seed = 42$. Los aumentos cromáticos se limitaron a variaciones moderadas de tono ($hsv_h = 0.05$), saturación ($hsv_s = 0.40$) y brillo ($hsv_v = 0.20$), complementadas con una perturbación de canal BGR de 0.10, configuración que aporta robustez ante variaciones de iluminación sin distorsionar la firma espectral del indicador de cúrcuma. Los aumentos geométricos se mantuvieron conservadores (rotación máxima de $\pm 10^\circ$, volteo horizontal con probabilidad 0.50 y volteo vertical con probabilidad 0.20) para preservar la coherencia física de la adquisición en laboratorio. Los umbrales de inferencia, determinados mediante barridos sobre los modelos candidatos en el bloque 6 del plan experimental, se fijaron en $conf = 0.50$ e $IoU = 0.55$ como la combinación que maximiza el equilibrio entre precisión y sensibilidad en el dominio evaluado.

4.2. Segmentación geométrica y fusión semántica por fragmento

En paralelo a la inferencia semántica del modelo YOLO, se ejecuta una segmentación geométrica 2D orientada a identificar los fragmentos físicos presentes en la escena. La imagen RGB se convierte a escala de grises, se suaviza mediante un filtro gaussiano de 5×5 píxeles y se umbraliza automáticamente mediante el método de Otsu. La máscara binaria resultante se regulariza con operaciones morfológicas de apertura y cierre sobre un elemento estructurante elíptico de 3×3 píxeles (1 iteración). A partir de los contornos externos se filtran componentes espurias mediante un umbral de área mínima equivalente al 0.1 % del área de imagen, con un mínimo de 10 píxeles. El esquema completo del pipeline, desde la imagen de entrada hasta la decisión por fragmento, se resume en la Figura 3. La fusión semántico-geométrica se obtiene mediante la intersección entre la máscara de cada fragmento y las máscaras semánticas generadas por YOLO.

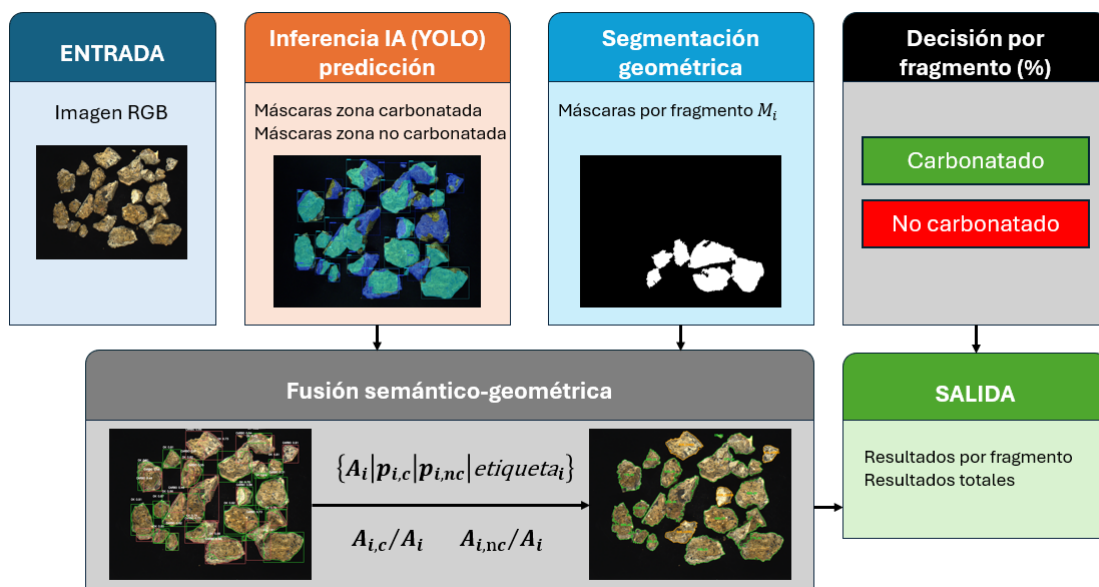


Figura 3. Esquema del pipeline de inferencia y fusión semántico-geométrica para la clasificación por fragmento.

Para cada fragmento i con área total A_i se calculan los solapamientos con las clases OK ($A_{i,nc}$) y CARBO ($A_{i,c}$), obteniendo las proporciones normalizadas $p_{i,nc} = A_{i,nc}/A_i$ y $p_{i,c} = A_{i,c}/A_i$. La etiqueta final se asigna mediante criterio de dominancia de área, el fragmento recibe la etiqueta de la clase cuya proporción es mayor, junto con el porcentaje dominante asociado.

Este criterio es coherente con la naturaleza espacial del problema y permite generar un registro individual con etiqueta, porcentaje dominante y estado de validez, habilitando la auditoría de cada predicción. El valor del porcentaje dominante aporta información adicional sobre la homogeneidad de la respuesta cromática dentro de cada fragmento. Valores elevados indican zonas de contraste nítido y clasificación robusta, mientras que valores próximos al umbral de paridad señalan fragmentos con distribución mixta, típicamente asociados a superficies de transición entre zonas carbonatadas y no carbonatadas o a áridos con escasa penetración del indicador. Estos casos no se descartan automáticamente, sino que se registran con su proporción exacta para que etapas posteriores del flujo de valorización puedan aplicar criterios de umbralización específicos según los requisitos del proceso. El conjunto de salidas por escena (etiquetas, proporciones por fragmento y métricas agregadas de confianza y recuento por clase) se almacena en MLflow como artefacto de cada ejecución de inferencia, garantizando la trazabilidad completa entre las condiciones de captura, los parámetros del modelo activo y los resultados de clasificación obtenidos. La Figura 4 muestra un ejemplo representativo de la campaña con el modelo v2 (optimizado), ilustrando las salidas del pipeline sobre una escena de 8 fragmentos H01 tratados con indicador de cúrcuma.

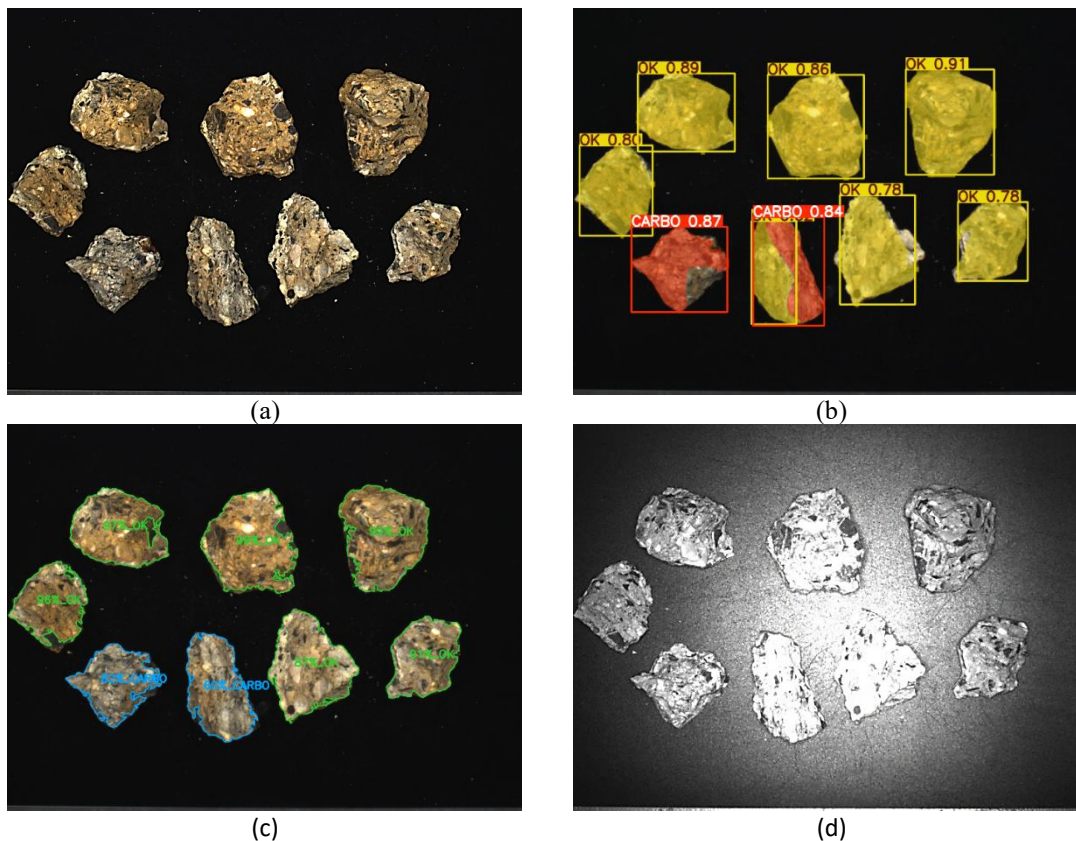


Figura 4. Imagen RGB capturada (a), fusión semántico-geométrica para la decisión por fragmento, máscaras YOLO (b) y resultado por elemento con etiqueta dominante y porcentaje (c), Imagen NIR para el control del contenido de humedad (d).

5 Resultados

La evaluación del sistema se estructura en 2 niveles complementarios. El análisis de las métricas de entrenamiento obtenidas durante la optimización del modelo y la validación funcional del sistema completo en condiciones de laboratorio. El primer nivel permite cuantificar el efecto de la ampliación del dataset sobre el rendimiento del modelo de segmentación, aislando la variable de tamaño y diversidad del conjunto de entrenamiento bajo condiciones de arquitectura e hiperparámetros constantes. El segundo nivel determina si las mejoras métricas observadas se traducen en un comportamiento operativo estable y reproducible cuando el sistema procesa escenas reales con indicador de cúrcuma aplicado. La validación funcional se estructuró en 3 ensayos complementarios. Los 2 primeros evaluaron el comportamiento nominal del sistema con el indicador de cúrcuma, mediante una campaña de referencia con $n = 5$ observaciones y una campaña de estabilidad temporal con $n = 21$ observaciones. El tercer ensayo se planteó como control fuera de dominio, empleando fenolftaleína como reactivo cromáticamente distinto al indicador usado en el entrenamiento, con $n = 5$ observaciones. Esta organización permite separar la estabilidad del sistema en condiciones nominales de su respuesta ante señales cromáticas no aprendidas. La Tabla 2 compara las métricas finales de los modelos v1 (230 imágenes) y v2 (308 imágenes), bajo condiciones estrictamente comparables, misma arquitectura, mismos hiperparámetros e idénticos umbrales de inferencia. La única diferencia entre ambas configuraciones es el tamaño y la diversidad del conjunto de entrenamiento.

Tabla 2. Métricas operativas de las campañas nominales con indicador de cúrcuma.

Métrica	Modelo v1	Modelo v2	Variación
mAP50-95 (B)	0.420	0.430	+0.010
mAP50-95 (M)	0.370	0.380	+0.010
mAP50 (B)	0.693	0.681	-0.012
mAP50 (M)	0.689	0.688	≈ 0
Precisión (B)	0.700	0.670	-0.030
Precisión (M)	0.712	0.690	-0.022
Recall (B)	0.676	0.681	+0.005
Recall (M)	0.678	0.677	≈ 0
Pérdida segmentación	2.50	2.31	-0.19

Los indicadores más exigentes en términos de fidelidad espacial de los contornos, mAP50-95 para cajas (B) y máscaras (M), presentan una mejora consistente en el modelo v2 mAP50-95(B) pasa de 0.420 a 0.430 y mAP50-95(M) de 0.370 a 0.380. Dado que esta métrica integra umbrales de IoU entre 0.50 y 0.95, penaliza de forma más severa las discrepancias geométricas en los contornos y constituye el indicador más representativo de calidad espacial efectiva para un sistema orientado a clasificación por fragmento. La pérdida de segmentación se redujo de 2.50 a 2.31, indicando mayor consistencia geométrica de las máscaras generadas por el modelo v2. La reducción observada en mAP50(B) (de 0.693 a 0.681) y en precisión (de 0.700 a 0.670 en cajas y de 0.712 a 0.690 en máscaras) es coherente con un modelo expuesto a mayor variabilidad visual al incorporar imágenes más diversas, el modelo reduce su ajuste a patrones específicos del conjunto inicial y desplaza el equilibrio hacia mayor sensibilidad en condiciones heterogéneas. El recall se mantiene estable o mejora ligeramente $0.676 \rightarrow 0.681$ en cajas y $0.678 \rightarrow 0.677$ en máscaras.

En el contexto del sistema experto, la sensibilidad resulta especialmente relevante, ya que los falsos negativos (no detectar zonas carbonatadas reales) tienen mayor impacto operativo que una clasificación conservadora hacia CARBO. La validación funcional del prototipo se organizó en 3 campañas. La Tabla 3 sintetiza las métricas operativas registradas en MLflow durante la campaña de referencia ($n = 5$ observaciones) y la campaña de estabilidad ($n = 21$ observaciones), ambas con indicador de cúrcuma bajo condiciones nominales de laboratorio. La estabilidad de los recuentos por clase en la campaña de $n=21$ se atribuye al uso del mismo conjunto de fragmentos con disposición controlada, lo que elimina la variabilidad inter-escena presente en la campaña de referencia.

Tabla 3. Métricas operativas de las campañas de referencia y estabilidad (media $\pm \sigma$, CV %).

Métrica operativa	Valor (ref. / estabilidad)	CV (%)
Detecciones totales (media $\pm \sigma$)	12.00 \pm 0.00 / 9.00 \pm 0.01	0.00 / 0.07
Confianza media (media $\pm \sigma$)	0.8287 \pm 0.0029 / 0.8109 \pm 0.0019	0.36 / 0.24
Fragmentos OK por escena (media $\pm \sigma$)	6.85 \pm 0.65 / 6.00 \pm 0.00	9.55 / 0.00
Fragmentos CARBO por escena (media $\pm \sigma$)	1.30 \pm 0.46 / 2.00 \pm 0.00	35.25 / 0.00
Dominancia media por fragmento (media $\pm \sigma$)	79.62 \pm 3.31 / 89.63 \pm 0.16	4.15 / 0.18

En la campaña de referencia, las métricas globales de detección y confianza muestran alta estabilidad (detecciones: CV = 0.00 % y confianza media: CV = 0.36 %). La variabilidad más elevada se presenta en el recuento de fragmentos por clase, especialmente en `seg_num_carbo` (CV = 35.25 %), lo que indica que algunos fragmentos en la frontera de dominancia alternan su clasificación entre registros consecutivos. En la campaña de estabilidad, realizada con una ventana de observación más amplia ($n = 21$), los recuentos convergen a valores invariantes (`seg_num_ok` = 6.00 \pm 0.00, CV = 0 %, `seg_num_carbo` = 2.00 \pm 0.00, CV = 0 %) y la dominancia media mejora tanto en magnitud como en dispersión (89.63 \pm 0.16, CV = 0.18 %). Este resultado confirma que la clasificación por fragmento se mantiene estable cuando la escena y el indicador son compatibles con las condiciones de captura del dataset y la ventana de observación es suficientemente representativa. El control de negativos, realizado con fragmentos tratados con fenolftaleína como señal cromática fuera del rango de imágenes con el que fue entrenado, produjo un desplazamiento del balance OK/CARBO respecto a la campaña nominal (`seg_num_ok`: 4.40 \pm 0.80 y `seg_num_carbo`: 4.20 \pm 0.40), sin refuerzo espurio de la clase OK ante colores no aprendidos. Ninguno de los 5 registros reforzó la clase OK ante el color magenta de la fenolftaleína, lo que descarta que el modelo trate cualquier zona coloreada como zona no carbonatada.

5.1. Discusión

La mejora de mAP50-95 al ampliar el dataset es coherente con un escenario donde la arquitectura y los hiperparámetros se mantuvieron constantes, la ganancia procede de la diversificación del conjunto de entrenamiento, no de cambios arquitectónicos. Frente a los métodos basados en umbralización cromática, el modelo profundo integra textura, forma y contexto local, la segmentación geométrica clásica complementa esta capacidad al asociar las predicciones a elementos físicos individuales, resolviendo ambos problemas de forma desacoplada y adecuada a la heterogeneidad de los residuos de hormigón. La salida del sistema (porcentaje OK o CARBO por fragmento) expresa dominancia de área proyectada en imagen, no

una magnitud química absoluta ni una certificación normativa, por lo que su integración en flujos de valorización requiere calibración frente a ensayos de referencia normalizados. Las principales limitaciones se relacionan con el dominio de entrenamiento, un único tipo de hormigón en condiciones de laboratorio controladas. La transferencia a entornos industriales exige ampliar el dataset a nuevas tipologías de hormigón, diversificar condiciones de iluminación e incorporar calibración radiométrica periódica.

6 Conclusiones y trabajo futuro

Se ha desarrollado y validado a nivel de laboratorio un sistema experto de visión artificial para la identificación automática de zonas carbonatadas y no carbonatadas en fragmentos de hormigón de demolición, empleando un indicador de pH inocuo basado en cúrcuma y un modelo YOLO11n-seg de segmentación de instancias. La fusión entre las máscaras semánticas del modelo y la segmentación geométrica por fragmento permite obtener una etiqueta de dominancia y un porcentaje asociado para cada elemento físico, de forma objetiva, reproducible y trazable. La ampliación del dataset de 230 a 308 imágenes produjo mejoras en los indicadores más exigentes de calidad geométrica (mAP50-95) y una reducción de la pérdida de segmentación, manteniendo el recall y asumiendo una reducción moderada en precisión coherente con la mayor diversidad del conjunto de entrenamiento. La validación funcional confirmó estabilidad bajo condiciones nominales de laboratorio, con recuentos invariantes por clase en la campaña de estabilidad y comportamiento conservador ante señales cromáticas fuera del dominio de entrenamiento. La metodología propuesta constituye una base reproducible para avanzar desde la inspección colorimétrica manual hacia la clasificación automatizada por fragmento. Como trabajo futuro, se requiere ampliar el dominio experimental a nuevas tipologías de hormigón, diversificar condiciones de iluminación, definir umbrales trazables frente a ensayos de referencia normalizados y validar el sistema en flujos de medida con mayor variabilidad ambiental.

7 Agradecimientos

El presente trabajo se ha desarrollado en el marco del proyecto CO2MCRETE, expediente MIG-20241096, financiado por el Centro para el Desarrollo Tecnológico y la Innovación (CDTI) a través del Programa de Misiones.

8 Referencias

- [1] A. Leemann and F. Moro, 'Carbonation of concrete: the role of CO₂ concentration, relative humidity and CO₂ buffer capacity', *Mater. Struct.*, vol. 50, no. 1, p. 30, Feb. 2017, doi: 10.1617/s11527-016-0917-2.
- [2] I. Nunez and M. L. Nehdi, 'Machine learning prediction of carbonation depth in recycled aggregate concrete incorporating SCMs', *Constr. Build. Mater.*, vol. 287, p. 123027, Jun. 2021, doi: 10.1016/j.conbuildmat.2021.123027.
- [3] E. M. Golafshani, A. Behnood, T. Kim, T. Ngo, and A. Kashani, 'A framework for low-carbon mix design of recycled aggregate concrete with supplementary cementitious materials using machine learning and optimization algorithms', *Structures*, vol. 61, p. 106143, Mar. 2024, doi: 10.1016/j.istruc.2024.106143.
- [4] N. Vogler, M. Lindemann, P. Drabetzki, and H.-C. Kühne, 'Alternative pH-indicators for determination of carbonation depth on cement-based concretes', *Cem. Concr. Compos.*, vol. 109, p. 103565, May 2020, doi: 10.1016/j.cemconcomp.2020.103565.
- [5] S. C. Yepes, 'Indicadores de pH inocuos sustitutos de la fenolftaleína'. *Revista Técnica Cemento Hormigón*, (975), 10–13, 2020.

- [6] S. Chinchón-Payá, 'Uso de antocianinas como indicadores'. *Revista Técnica Cemento Hormigón*, (1005), 14–17. ISSN: 0008-8919, 2022.
- [7] J.-I. Choi, Y. Lee, Y. Y. Kim, and B. Y. Lee, 'Image-processing technique to detect carbonation regions of concrete sprayed with a phenolphthalein solution', *Constr. Build. Mater.*, vol. 154, pp. 451–461, Nov. 2017, doi: 10.1016/j.conbuildmat.2017.07.205.
- [8] N. Giulietti *et al.*, 'Automated measurement system for detecting carbonation depth: Image-processing based technique applied to concrete sprayed with phenolphthalein', *Measurement*, vol. 175, p. 109142, Apr. 2021, doi: 10.1016/j.measurement.2021.109142.
- [9] R. G. D. Silveira and G. C. Isaia, 'Proposal of a digital image analysis method for determining the carbonated front in concretes', *Ambiente Construído*, vol. 22, no. 3, pp. 275–298, Sep. 2022, doi: 10.1590/s1678-86212022000300619.
- [10] K. Zhang, M. Yio, H. Wong, and N. Buenfeld, 'Development of more accurate methods for determining carbonation depth in cement-based materials', *Cem. Concr. Res.*, vol. 175, p. 107358, Jan. 2024, doi: 10.1016/j.cemconres.2023.107358.
- [11] X. Wang, Q. Yang, X. Peng, and F. Qin, 'A Review of Concrete Carbonation Depth Evaluation Models', *Coatings*, vol. 14, no. 4, p. 386, Mar. 2024, doi: 10.3390/coatings14040386.
- [12] Z. Liu, Y. Cao, Y. Wang, and W. Wang, 'Computer vision-based concrete crack detection using U-net fully convolutional networks', *Autom. Constr.*, vol. 104, pp. 129–139, Aug. 2019, doi: 10.1016/j.autcon.2019.04.005.
- [13] Y. Ji, H. Zhang, Z. Zhang, and M. Liu, 'CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances', *Inf. Sci.*, vol. 546, pp. 835–857, Feb. 2021, doi: 10.1016/j.ins.2020.09.003.
- [14] M. Sohaib, M. Arif, and J.-M. Kim, 'Evaluating YOLO Models for Efficient Crack Detection in Concrete Structures Using Transfer Learning', *Buildings*, vol. 14, no. 12, p. 3928, Dec. 2024, doi: 10.3390/buildings14123928.
- [15] J. A. Guzmán-Torres, F. J. Domínguez-Mota, W. Martínez-Molina, M. Z. Naser, G. Tinoco-Guerrero, and J. G. Tinoco-Ruíz, 'Damage detection on steel-reinforced concrete produced by corrosion via YOLOv3: A detailed guide', *Front. Built Environ.*, vol. 9, p. 1144606, Mar. 2023, doi: 10.3389/fbuil.2023.1144606.
- [16] M. A. R. Alif and M. Hussain, 'YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain', Jun. 14, 2024, *arXiv*: arXiv:2406.10139. doi: 10.48550/arXiv.2406.10139.
- [17] M. L. Ali and Z. Zhang, 'The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection', *Computers*, vol. 13, no. 12, p. 336, Dec. 2024, doi: 10.3390/computers13120336.
- [18] Z. Qiu, X. Huang, Z. Sun, S. Li, and J. Wang, 'GS-YOLO-Seg: A Lightweight Instance Segmentation Method for Low-Grade Graphite Ore Sorting Based on Improved YOLO11-Seg', *Sustainability*, vol. 17, no. 12, p. 5663, Jun. 2025, doi: 10.3390/su17125663.
- [19] D. Kreuzberger, N. Köhl, and S. Hirschl, 'Machine Learning Operations (MLOps): Overview, Definition, and Architecture', *IEEE Access*, vol. 11, pp. 31866–31879, 2023, doi: 10.1109/ACCESS.2023.3262138.
- [20] N. Hewage and D. Meedeniya, 'Machine Learning Operations: A Survey on MLOps Tool Support', 2022, doi: 10.48550/arXiv.2202.10169.
- [21] J. Valença, L. M. S. Gonçalves, and E. Júlio, 'Damage assessment on concrete surfaces using multi-spectral image analysis', *Constr. Build. Mater.*, vol. 40, pp. 971–981, Mar. 2013, doi: 10.1016/j.conbuildmat.2012.11.061.
- [22] T. Honda, D. Sugimura, and T. Hamamoto, 'Multi-Frame RGB/NIR Imaging for Low-Light Color Image Super-Resolution', *IEEE Trans. Comput. Imaging*, vol. 6, pp. 248–262, 2020, doi: 10.1109/TCI.2019.2948779.
- [23] H. Zheng *et al.*, 'Early season detection of rice plants using RGB, NIR-G-B and multispectral images from unmanned aerial vehicle (UAV)', *Comput. Electron. Agric.*, vol. 169, p. 105223, Feb. 2020, doi: 10.1016/j.compag.2020.105223.
- [24] N. Lauchande, *Machine learning engineering with MLflow: manage the end-to-end machine learning lifecycle with MLflow*. Birmingham Mumbai: Packt, 2021.