



**Pontificia Universidad
Católica del Ecuador**
Seréis mis testigos

MANABÍ

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

SEDE MANABÍ

CARRERA DE INGENIERÍA CIVIL

TRABAJO DE TITULACIÓN

RED NEURONAL CONVOLUCIONAL PARA LA DETECCIÓN DE FISURAS EN

ESTRUCTURAS DE HORMIGÓN

LÍNEA DE INVESTIGACIÓN

HÁBITAT, INFRAESTRUCTURA Y MOVILIDAD

SUBLÍNEA DE INVESTIGACIÓN

CONSTRUCCIONES CIVILES

PREVIO AL TÍTULO DE

INGENIERA CIVIL

AUTOR(A)

MELISSA MARÍA CEDENO FARÍAS

TUTOR(A)

MGTR. MAURICIO HENRY COLPARI POZZO

PORTOVIEJO, SEPTIEMBRE 2025

Certificado del trabajo de integración curricular

De: Mgtr, Mauricio Henry Colpari Pozzo

Tutor trabajo de integración curricular

Para: Mgtr, Yandri Xavier Vélez Molina

Coordinador de la carrera de Ingeniería Civil

CERTIFICA

En mi calidad de tutor del trabajo de integración curricular, certifico haber revisado el presente manuscrito de investigación, el mismo que se ajusta a las normas vigentes de la Pontificia Universidad Católica del Ecuador Sede Manabí, cumpliendo la Normativa del Trabajo de Integración Curricular; en consecuencia, es apto para su presentación y sustentación

(F)_____

Mgtr., Mauricio Henry Colpari Pozzo

C.I 0965673726

Acta de aprobación del tribunal

El jurado examinador aprueba el presente trabajo de integración curricular en nombre de la Pontificia Universidad Católica del Ecuador, Sede Manabí

(F) _____

Mgtr., Mauricio Henry Colpari Pozzo

Primer lector

(F) _____

Mgtr., María Tatiana Ordóñez Zambrano

Segundo lector

(F) _____

Mgtr., José Ramón Alarcón Loor

Tercer lector

Declaración de Originalidad

Este manuscrito no contiene ningún tipo de material que ha sido aceptado para la obtención de un título universitario en otra institución, excepto en forma de información de soporte que ha sido debidamente citada en mi trabajo.

Este trabajo es de total responsabilidad del autor, quien declara bajo juramento que ninguna sección de este trabajo de integración curricular infringe los derechos de autor de nadie

(F)_____

Melissa María Cedeño Farías

C.I. 1315588127

Dirección: Urb. San Marino, Av. Manabí

Correo electrónico: melicefa25@hotmail.com

Celular: 0985605011

Declaración de derechos de autor

Autorizo a la Pontificia Universidad Católica del Ecuador a distribuir este manuscrito de investigación en medios físicos y electrónicos con el fin de promover la divulgación de mis resultados a la comunidad científica y a la sociedad en general. Adicionalmente autorizo el uso de los contenidos de esta investigación como bibliografía para fines académicos, por cualquier medio o procedimiento, citando como fuente de información al autor de este trabajo

(F)_____

Melissa María Cedeño Farías

C.I. 1315588127

Dedicatoria

A mi mamá, por ser la inspiración constante en mi vida, por su amor incondicional, sus sacrificios y su ejemplo de fortaleza. Este logro es tan suyo como mío.

Agradecimientos

Quiero expresar mi más sincero agradecimiento al Mgtr. Mauricio Colpari, tutor de esta tesis y miembro del grupo de investigación, por su constante apoyo, paciencia y valiosas orientaciones que guiaron este trabajo desde sus inicios. De igual manera, agradezco a la Mgtr. Tatiana Ordóñez por sus aportes y acompañamiento durante el desarrollo del proyecto.

Extiendo también mi gratitud al Dr. Jesús Chavarría, cuya guía y experiencia fueron fundamentales para enriquecer este proceso de investigación.

Agradezco a mi universidad y a todos los docentes que contribuyeron con su conocimiento a lo largo de mi formación académica, así como a mis compañeros de carrera y miembros del grupo de investigación, por el apoyo mutuo y el trabajo en equipo que hicieron posible alcanzar esta meta.

De manera especial, agradezco a mi mamá, Gina Farías, por su amor incondicional, motivación y ejemplo de fortaleza que me impulsaron a culminar este proceso, y a mi papá, Roosevelt Cedeño, por su apoyo constante que me acompañó en todo momento. Asimismo, a mis tías, Luisa Farías y Fanny Farías, quienes con su apoyo me brindaron la oportunidad de seguir adelante y mantenerme firme en este camino.

Finalmente, expreso mi gratitud a toda mi familia y amigos, quienes con sus palabras de aliento y compañía me dieron la fuerza necesaria para culminar esta etapa tan importante de mi vida.

Resumen

La evaluación automatizada de fisuras en estructuras de hormigón representa una necesidad creciente en la ingeniería civil. La identificación temprana de estas patologías es fundamental para garantizar la seguridad y durabilidad de las estructuras. Los métodos tradicionales, basados en inspecciones visuales, suelen ser costosos, subjetivos y poco eficientes. Esta investigación propone un enfoque basado en aprendizaje profundo para abordar estas limitaciones, evaluando el desempeño de dos arquitecturas de redes neuronales convolucionales (CNN) en la detección de fisuras. Los modelos VGG16 y ResNet50 fueron entrenados y validados utilizando una base de datos original de 8,000 imágenes. Las métricas de rendimiento mostraron que ResNet50 alcanzó 90.80% de exactitud, 86.93% de precisión y 90.62% de sensibilidad, mientras que VGG16 obtuvo 88.95% de exactitud, 96.75% de precisión y 84.42% de sensibilidad. Estos resultados destacan el potencial de las CNN como herramientas para fortalecer el monitoreo estructural, promoviendo sistemas de inspección más ágiles, objetivos y adaptables.

Palabras clave: red neuronal convolucional, aprendizaje profundo, visión por computadora, patologías del hormigón, inspección estructural.

Abstract

Automated crack assessment in concrete structures is becoming an increasingly critical task in civil engineering. The early detection of such pathologies is essential to ensure the safety, serviceability, and long-term durability of reinforced concrete structures. Conventional approaches, typically relying on visual inspections, are often costly, subjective, and inefficient. To address these limitations, this study introduces a deep learning-based methodology that evaluates the performance of two convolutional neural network (CNN) architectures for crack detection. The VGG16 and ResNet50 models were trained and validated on a novel dataset comprising 8,000 images. Experimental results indicate that ResNet50 achieved an accuracy of 90.80%, a precision of 86.93%, and a recall of 90.62%, whereas VGG16 obtained an accuracy of 88.95%, a precision of 96.75%, and a recall of 84.42%. These findings demonstrate the effectiveness of CNNs as robust tools for structural health monitoring, fostering the development of more agile, objective, and adaptable inspection systems.

Keywords: convolutional neural network, deep learning, computer vision, concrete pathologies, structural inspection.

Tabla de Contenidos

Certificado del trabajo de integración curricular	i
Acta de aprobación del tribunal	ii
Declaración de Originalidad	iii
Declaración de derechos de autor	iv
Dedicatoria	v
Agradecimientos	vi
Resumen.....	vii
Abstract	viii
Tabla de Contenidos	ix
Índice de Tablas	xi
Índice de Figuras.....	xii
Índice de Ecuaciones.....	xiii
Introducción	1
Objetivos	4
Objetivo General	4
Objetivos Específicos.....	4
Metodología	5
Tipo de Investigación.....	5
Método General	6
<i>Método Basado en Redes Neuronales Convolucionales.....</i>	<i>6</i>

<i>Justificación del Enfoque Automatizado para Evaluación Estructural</i>	6
Fases del Estudio.....	7
<i>Fase 1: Recolección de Datos</i>	7
<i>Fase 2: Preprocesamiento de Imágenes</i>	11
<i>Fase 3: Diseño y Entrenamiento del Modelo CNN</i>	12
<i>Fase 4: Validación del Modelo (Aplicación en Campo)</i>	20
Materiales.....	22
Población y Muestra	24
<i>Descripción de la Obra Seleccionada</i>	24
<i>Número de las Imágenes</i>	26
Resultados	28
Levantamiento de Datos	28
Entrenamiento de CNN.....	29
<i>ResNet50</i>	29
<i>VGG16</i>	32
<i>Comparación de Resultados de CNN</i>	36
Discusión.....	38
Comparación Metodológica con Investigaciones Previas	39
Comparación del Desempeño de Métricas de Modelos.....	42
Conclusiones	48
Referencias Bibliográficas	51

Índice de Tablas

Tabla 1 Guía para Anchos Razonables de Fisuras en Hormigón Armado Bajo Cargas de Servicio	10
Tabla 2 Análisis Bibliográfico de Cantidad de Imágenes y Rendimiento de Modelos	26
Tabla 3 Resultados de la Captura y Procesamiento de Imágenes para la Base de Datos CNN	28
Tabla 4 Resultados Comparativos de los Modelos VGG16 y ResNet50.....	36

Índice de Figuras

Figura 1 <i>Proceso de Levantamiento de Datos</i>	8
Figura 2 <i>Muestras Representativas de Imágenes con Fisuras Capturadas</i>	9
Figura 3 <i>Medición del Ancho de Fisura en Acera de Hormigón Mediante Fisurómetro</i>	11
Figura 4 <i>Ejemplos de Imágenes Recortadas Utilizadas en el Entrenamiento de los Modelos CNN</i>	12
Figura 5 <i>Arquitectura Básica de VGG16</i>	15
Figura 6 <i>Arquitectura Básica de ResNet50</i>	17
Figura 7 <i>Proceso de Levantamiento de Datos para Conjunto de Validación</i>	20
Figura 8 <i>Medición de fisura en Elemento Estructural del Puente del Redondel San Juan</i> ...	21
Figura 9 <i>Cámara Sony Alpha-6100</i>	22
Figura 10 <i>Entorno de Desarrollo PyCharm 2025.1.1.1</i>	23
Figura 11 <i>Computadora Portátil Lenovo IdeaPad 3</i>	23
Figura 12 <i>Dron DJI Phantom 4 RTK</i>	24
Figura 13 <i>Puente Redondel San Juan</i>	25
Figura 14 <i>Evolución de la Exactitud de ResNet50</i>	30
Figura 15 <i>Evolución de la Precisión y Sensibilidad de ResNet50</i>	31
Figura 16 <i>Evolución de la Pérdida de ResNet50</i>	32
Figura 17 <i>Evolución de la Exactitud de VGG16</i>	33
Figura 18 <i>Evolución de la Precisión y Sensibilidad de VGG16</i>	34
Figura 19 <i>Evolución de la Pérdida de VGG16</i>	35
Figura 20 <i>Comparación del Desempeño entre VGG16 y ResNet50 en el Conjunto De Validación</i>	37

Índice de Ecuaciones

Ecuación 1 <i>Exactitud (Accuracy)</i>	18
Ecuación 2 <i>Precisión (Precisión)</i>	18
Ecuación 3 <i>Sensibilidad (Recall)</i>	18

Introducción

La inspección y el mantenimiento preventivo de las estructuras de hormigón son fundamentales para garantizar su durabilidad (Gjorv, 2011). Tanto en estructuras nuevas como en aquellas en servicio, la evaluación puede realizarse mediante ensayos no destructivos, los cuales permiten estimar la resistencia in situ del hormigón y detectar otras condiciones relevantes, como la presencia de vacíos internos o signos de corrosión activa en las armaduras (American Concrete Institute, 2023). Según Dyer (2014), el hormigón está expuesto a múltiples mecanismos de degradación que pueden afectar su desempeño estructural, entre ellos la contracción, la fisuración térmica, los ciclos de hielo-deshielo, la abrasión, la erosión y el ataque de sulfatos. Frente a estos factores de deterioro, el monitoreo de la salud estructural se vuelve una herramienta clave para evaluar el estado funcional de las estructuras y detectar posibles daños (J. Deng et al., 2022).

El análisis a gran escala de estructuras de hormigón es un proceso que demanda un alto consumo de tiempo y recursos (Koch et al., 2014). La inteligencia artificial proporciona nuevas herramientas, destacándose el uso de técnicas de aprendizaje automático, en particular las redes neuronales convolucionales (CNN, por sus siglas en inglés), para la detección de fisuras a partir del análisis de imágenes. Las CNN son un tipo de red neuronal de propagación hacia adelante con una arquitectura profunda, compuesta por capas de convolución, capas de activación, capas de agrupamiento, capas completamente conectadas y capas de abandono, las cuales trabajan en conjunto para extraer y procesar características relevantes en el análisis de datos (Wen et al., 2025). La detección optimizada de fisuras mediante redes neuronales convolucionales permite identificar de manera automática la presencia de fisuras en una imagen y señalar su ubicación con precisión, lo cual contribuye significativamente a la eficiencia del monitoreo estructural (K. Chen et al., 2019).

Asimismo, se han implementado tecnologías como vehículos aéreos no tripulados (VANT) para la inspección de estructuras como puentes de hormigón, integrando modelos de aprendizaje profundo para el procesamiento y análisis automatizado de los datos capturados (Zoubir et al., 2024). Además del uso predominante de imágenes 2D para el reconocimiento de defectos mediante aprendizaje profundo, se han implementado modelos avanzados de redes neuronales convolucionales 3D, como 3D InspectionNet, con el propósito de extraer características de defectos a partir de conjuntos de datos tridimensionales generados artificialmente (Shafiei Dizaji & Harris, 2019).

La aplicación de CNN para la detección de fisuras a través del análisis de imágenes no se restringe únicamente a estructuras de hormigón, sino que también puede emplearse en estructuras metálicas. En estudios previos, esta tecnología ha sido utilizada para analizar imágenes de las juntas de soldadura en puentes de acero de gran longitud (Quqa et al., 2022).

Las estructuras de hormigón, presentes en diversas aplicaciones como rascacielos, puentes y presas, experimentan un deterioro progresivo a lo largo de su vida útil. Esta degradación estructural hace que la detección en tiempo real de distintos tipos de daños sea una necesidad fundamental (Fu et al., 2020). El deterioro del hormigón puede generar pérdida de superficie del hormigón, fisuración o una combinación de ambos efectos. La presencia de fisuras puede afectar el desempeño estructural, además de facilitar la infiltración de agentes químicos dañinos que pueden comprometer la integridad del hormigón y del acero de refuerzo (Dyer, 2014). La corrosión del acero de refuerzo en estructuras de hormigón que ocurre con el paso del tiempo provoca una expansión volumétrica que genera microfisuras, lo que a su vez disminuye la durabilidad estructural (Lee et al., 2017). La fisuración es un fenómeno intrínseco del hormigón, por lo que, aunque no puede eliminarse por completo, es posible minimizar su presencia, con el fin de reducir no solo el riesgo de corrosión del

refuerzo, sino también los efectos estéticos y la percepción de inseguridad en los usuarios (Calavera, 2005).

Frente a las limitaciones de la inspección convencional de estructuras de hormigón, esta investigación, mediante la implementación de técnicas avanzadas de aprendizaje profundo y visión por computadora, busca optimizar los procesos de monitoreo estructural. A partir de una base de datos de imágenes de fisuras en estructuras de hormigón, se entrenan múltiples modelos CNN, seleccionando para este estudio las arquitecturas VGG16 y ResNet50. Posteriormente, se realiza un análisis comparativo entre ambas arquitecturas de CNN con el objetivo de evaluar su precisión y desempeño en la detección automática de fisuras, considerando estudios previos que han comparado el rendimiento de modelos como ResNet50, VGG16 y VGG19 (Junior et al., 2024). Ambos modelos son implementados en un entorno real, específicamente en el Puente del Redondel San Juan, con el propósito de validar su eficacia en la identificación de fisuras bajo condiciones operativas reales.

En investigaciones recientes, se ha evaluado el rendimiento de diversos métodos de aprendizaje de automático en la detección de fisuras, determinando que las CNN presentan los mejores resultados, alcanzando niveles de precisión entre el 95 % y el 99 %, incluso en imágenes de baja resolución (Rashid et al., 2024). Además, las CNN han demostrado su capacidad para analizar imágenes del mundo real, diferenciando con precisión entre distintos tipos de daños, como grafitis, deterioro superficial o fugas, y permitiendo la identificación exacta de fisuras en estructuras de hormigón (Tello-Gil et al., 2025).

Objetivos

Objetivo General

Desarrollar una red neuronal convolucional para la detección de fisuras en estructuras de hormigón.

Objetivos Específicos

- Levantar los datos de campo que retroalimente el modelo de programación de la red neuronal convolucional.
- Entrenar el código de programación del modelo de la red neuronal convolucional.

Metodología

Tipo de Investigación

Esta investigación siguió un enfoque cuantitativo, que se basó en la recolección y el análisis estadístico de datos numéricos obtenidos a partir del entrenamiento y validación de modelos de redes neuronales convolucionales (CNN) para la detección automática de fisuras en estructuras de hormigón.

La investigación tuvo un carácter aplicado, ya que se busca resolver un problema práctico en el área de la ingeniería civil, siendo este, la automatización de la detección de fisuras en estructuras de hormigón. Este trabajo se enfocó en implementar un modelo funcional que pueda ser utilizado en inspecciones reales, reduciendo la subjetividad y el tiempo asociados a los métodos manuales tradicionales.

Dado que la aplicación de CNN para la detección de fisuras en hormigón sigue siendo un área en desarrollo, este estudio adopta un enfoque exploratorio. Se investigaron diferentes arquitecturas de redes neuronales y técnicas de preprocesamiento para determinar cuáles ofrecen el mejor rendimiento en un contexto específico.

La investigación fue experimental debido a que se implementó un procedimiento comparativo entre dos modelos de redes neuronales convolucionales para determinar cuál ofrece un mejor desempeño en la detección de fisuras en estructuras de hormigón. Este enfoque experimental permitió validar empíricamente la eficacia de cada arquitectura bajo condiciones similares.

Método General

Método Basado en Redes Neuronales Convolucionales

La red neuronal convolucional es una arquitectura de aprendizaje profundo de propagación hacia adelante, especialmente diseñada para el procesamiento de imágenes (Luo, 2021). Las CNN se inspiran en el funcionamiento de la corteza visual humana, para emular el procesamiento de la información visual (Guerri et al., 2024). Esto las hace particularmente eficaces en tareas que implican análisis y reconocimiento de patrones (Płotka et al., 2021). Las CNN destacan en el reconocimiento de bordes, formas y texturas en imágenes, por lo que son adecuadas para tareas de clasificación de imágenes y detección de objetos (Ramazhan et al., 2025). Avances recientes en el aprendizaje profundo han hecho que los métodos CNN alcancen un éxito notable en la detección de defectos, atribuido a su sólida percepción y sus capacidades de ajuste no lineal (Deng et al., 2020).

Justificación del Enfoque Automatizado para Evaluación Estructural

La necesidad de implementar una evaluación estructural automatizada surge de las limitaciones inherentes de las metodologías tradicionales, que generalmente requieren inspecciones y análisis manuales, caracterizados por ser procesos intensivos en recursos, lentos y susceptibles a errores humanos y sesgos subjetivos (Aslan et al., 2019). Los sistemas automatizados y la visión artificial ofrecen una alternativa eficiente, capaz de mitigar dichas limitaciones y optimizar tanto la inspección como el mantenimiento preventivo (Modarres et al., 2018). Las CNN pueden ser diseñadas para identificar diversas patologías estructurales, como fisuras, baches y otros tipos de defectos, en diferentes tipos de superficies (Yabi et al., 2025).

Las fisuras, como uno de los defectos más comunes en estructuras de hormigón, pueden reflejar problemas graves en la integridad estructural, afectando la seguridad tanto de

la estructura como de sus usuarios (Wu et al., 2025). Por ello, la detección precisa, rápida y automatizada de fisuras en superficies estructurales constituye un aspecto fundamental para múltiples aplicaciones en ingeniería civil, constituyendo además la etapa inicial en diversos procesos de diagnóstico e inspección (Ai et al., 2023). Las CNN permiten el monitoreo estructural automatizado para la detección temprana de fisuras en tiempo real, facilitando intervenciones preventivas y reduciendo el riesgo de fallas catastróficas mediante alertas oportunas (Sargiotis, 2024).

Fases del Estudio

Fase 1: Recolección de Datos

La primera etapa de la investigación comprendió la adquisición sistemática de imágenes de estructuras de hormigón en condiciones reales de servicio. Para este propósito, se llevó a cabo un protocolo de captura fotográfica, representado en la Figura 1, mediante el cual se documentó una amplia variedad de fisuras en estructuras de hormigón ubicadas en el cantón Portoviejo, tomadas de diversas estructuras como vigas, columnas, aceras, bordillos, entre otros. El proceso de adquisición de imágenes se realizó utilizando la cámara Sony Alpha-6100 PZ 16-50mm F3.5-5.6 OSS, la cual permitió obtener fotografías de alta resolución con un nivel de detalle adecuado para el posterior análisis computacional.

Figura 1*Proceso de Levantamiento de Datos*

Nota. Trabajo de campo realizado durante la primera etapa de la investigación, enfocado en el levantamiento fotográfico de diversas estructuras de hormigón.

Se capturaron un total de 715 imágenes en exteriores, con una resolución de 6024×4024 píxeles, utilizando el formato RAW. Las fotografías se realizaron bajo condiciones de iluminación natural, empleando una configuración de cámara con sensibilidad ISO 400, apertura de diafragma $f/8$ y velocidad de obturación de $1/125$ segundos, con el fin de garantizar una adecuada nitidez, contraste y nivel de detalle para el análisis posterior. En la Figura 2 se presentan algunos ejemplos representativos de estas imágenes capturadas, que ilustran la variedad de superficies y condiciones observadas durante el levantamiento en campo.

Figura 2*Muestras Representativas de Imágenes con Fisuras Capturadas*

Nota. Imágenes con resolución de 6024×4024 en formato RAW, capturadas en exteriores bajo condiciones de iluminación natural.

Además de la captura fotográfica, se realizaron mediciones manuales del ancho de fisuras utilizando un fisurómetro, el cual es una herramienta fundamental para la evaluación visual de fisuras en elementos de hormigón. Este instrumento permite verificar la presencia de fisuración y medir con precisión su ancho, aspecto clave para verificar el cumplimiento con los límites máximos aceptables establecidos en la normativa ACI 224R-01 (ACI, 2001), los cuales varían según las condiciones de exposición de la estructura, como se muestra en la Tabla 1.

Tabla 1

Guía para Anchos Razonables de Fisuras en Hormigón Armado Bajo Cargas de Servicio

Condición de exposición	Ancho de fisura	
	in	mm
Aire seco o membrana protectora	0.016	0.41
Humedad, aire húmedo, suelo	0.012	0.30
Productos químicos para deshielo	0.007	0.18
Agua y rocío de mar, humectación y secado	0.006	0.15
Estructuras de retención de agua	0.004	0.10

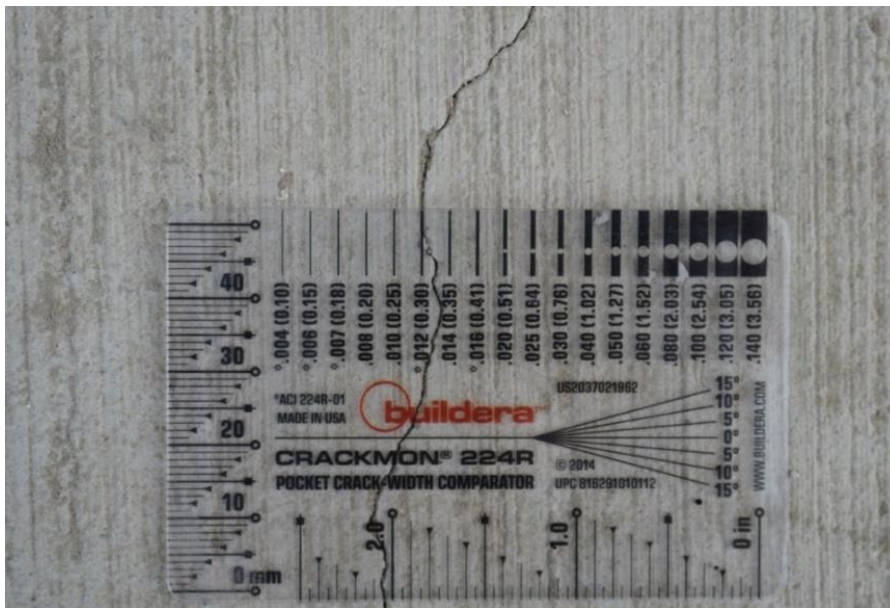
Nota. Adaptada de Control of Cracking in Concrete Structures por American Concrete Institute, 2001.

La medición del espesor de las fisuras es esencial en el diagnóstico estructural porque permite determinar en qué momento una fisura pasa de ser un defecto estético o superficial, a representar un riesgo estructural o de durabilidad. El ingreso de humedad, agentes químicos, cloruros o sulfatos a través de fisuras que superan los límites establecidos por la normativa puede acelerar la corrosión del acero de refuerzo, comprometiendo la vida útil de la estructura.

Este criterio normativo puede ser ilustrado mediante el análisis de imágenes tomadas en campo con la ayuda de un fisurómetro. Por ejemplo, en la Figura 3, correspondiente a una acera de hormigón, se observa una fisura de 0.30 mm. De acuerdo con la ACI, (2001), este valor representa el límite máximo tolerable para estructuras expuestas a humedad, aire húmedo o contacto con suelo, por lo que dicha fisura amerita monitoreo o intervención preventiva.

Figura 3

Medición del Ancho de Fisura en Acera de Hormigón Mediante Fisurómetro



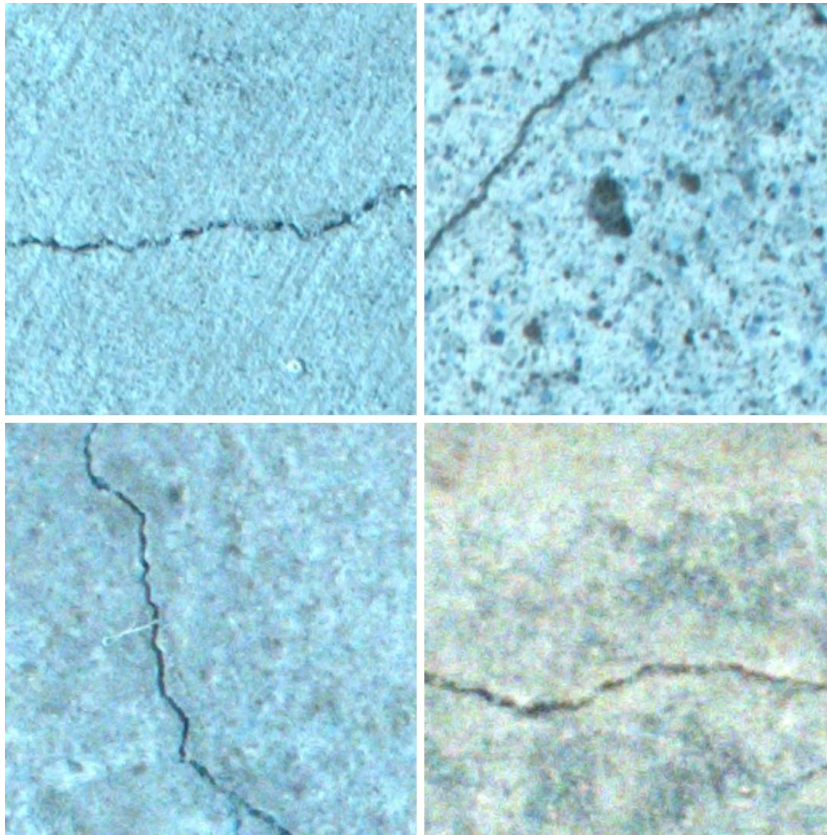
Nota. La imagen muestra una fisura de 0.30 mm detectada en una acera de hormigón, medida con un fisurómetro durante el levantamiento de datos en campo.

Fase 2: Preprocesamiento de Imágenes

Las 715 imágenes de 6024×4024 píxeles capturadas fueron segmentadas en subimágenes de 256×256 píxeles mediante el entorno de desarrollo PyCharm 2025.1.1.1. A partir de estos recortes, se seleccionaron manualmente 6,000 imágenes para conformar la base de datos de entrenamiento, descartando aquellas que presentaban desenfoque u otros defectos que comprometieran la calidad del análisis. La base de datos resultante se distribuyó en un 60 % de imágenes con presencia de fisuras y un 40 % sin fisuras. En la Figura 4 se muestran ejemplos de estas imágenes recortadas.

Figura 4

Ejemplos de Imágenes Recortadas Utilizadas en el Entrenamiento de los Modelos CNN



Nota. Cada subimagen de 256×256 píxeles fue extraída a partir de fotografías originales capturadas en campo.

Fase 3: Diseño y Entrenamiento del Modelo CNN

El diseño y entrenamiento de los modelos de CNN se realizó utilizando el lenguaje de programación Python, mediante el entorno de desarrollo PyCharm 2025.1.1.1. El proceso se ejecutó en una computadora portátil marca Lenovo. Durante el proceso, se emplearon las bibliotecas TensorFlow, Matplotlib y NumPy, para la implementación del modelo, la visualización de resultados y el manejo de estructuras de datos numéricas, respectivamente.

TensorFlow es una biblioteca de código abierto para aprendizaje profundo, desarrollada por Google en 2015, que permite a definición, entrenamiento e implementación de modelos de aprendizaje automático (Goldsborough, 2016). Esta herramienta emplea un

grafo de flujo de datos para representar de manera integral los cálculos y estados involucrados en un algoritmo de aprendizaje, incluyendo operaciones matemáticas, parámetros, reglas de actualización y procesos de preprocesamiento de datos (Abadi et al., 2016). En comparación con otras bibliotecas populares en el campo, TensorFlow ofrece un rendimiento computacional superior y una funcionalidad excelente (Prakash et al., 2021), lo que facilita significativamente tanto la investigación como la aplicación de modelos de redes neuronales (Pang et al., 2020).

Matplotlib es una biblioteca de gráficos y diagramas de Python que permite generar una variedad de tipos de gráficos o tablas en diversos formatos, entre los cuales están gráficos de líneas, de dispersión, de calor, de barras, circulares y 3D (Hunt, 2019). Se caracteriza por ofrecer un lenguaje conciso, alta precisión en el trazado y un código simple, lo que facilita su comprensión (Yang et al., 2021).

NumPy, abreviatura de *Numerical Python*, es una biblioteca fundamental de Python para el cálculo matemático y el procesamiento de arreglos (Rajagopalan, 2021). Está especialmente orientada a la computación científica y al análisis de datos, y constituye la base de numerosos paquetes matemáticos y científicos del ecosistema de Python (Nelli, 2015). A diferencia de las listas nativas de Python, que presentan limitaciones en velocidad, NumPy proporciona el objeto ndarray, una estructura eficiente para representar arreglos de n dimensiones, capaz de operar hasta 50 veces más rápido que las listas convencionales (Sapre & Vartak, 2020).

Los modelos de redes neuronales convolucionales seleccionados para esta investigación fueron VGG16 y ResNet50, debido a sus reconocidas capacidades para la clasificación de imágenes y a su uso frecuente en estudios recientes, como los de Junior et al. (2024), Mayya et al. (2024), Philip et al. (2023), Ojeda et al. (2023), Ali et al. (2021) y Xu et

al. (2019). ResNet50 ha demostrado un excelente rendimiento de generalización, presentando bajas tasas de error en tareas de reconocimiento de imágenes, lo que lo convierte en una herramienta eficaz para la clasificación de patrones visuales complejos (Konovalenko et al., 2020), como es el caso de las fisuras en estructuras de hormigón. Por su parte, VGG16 permite una reducción significativa del número de parámetros al simplificar las capas del modelo. Esta propiedad no solo mejora la eficiencia computacional, sino que también facilita la extracción de características relevantes con un menor costo de procesamiento (Ojeda et al., 2023).

En conjunto, ambos modelos ofrecen un equilibrio entre profundidad, capacidad de generalización y eficiencia, aspectos clave para evaluar su desempeño en un entorno real de inspección estructural automatizada.

VGG16. VGG16 es una red neuronal convolucional empleada para tareas de clasificación y detección de objetos, siendo ampliamente utilizada en el ámbito del reconocimiento de imágenes (Philip et al., 2023). Entre las distintas arquitecturas de CNNs, VGG16 se ha consolidado como una de las más utilizadas, cuya principal contribución radica en su topología homogénea. Esta arquitectura incrementa la profundidad de la red mediante el apilamiento secuencial de capas convolucionales y usa filtros convolucionales de tamaño reducido, específicamente de 2×2 y 3×3 (Kim et al., 2021).

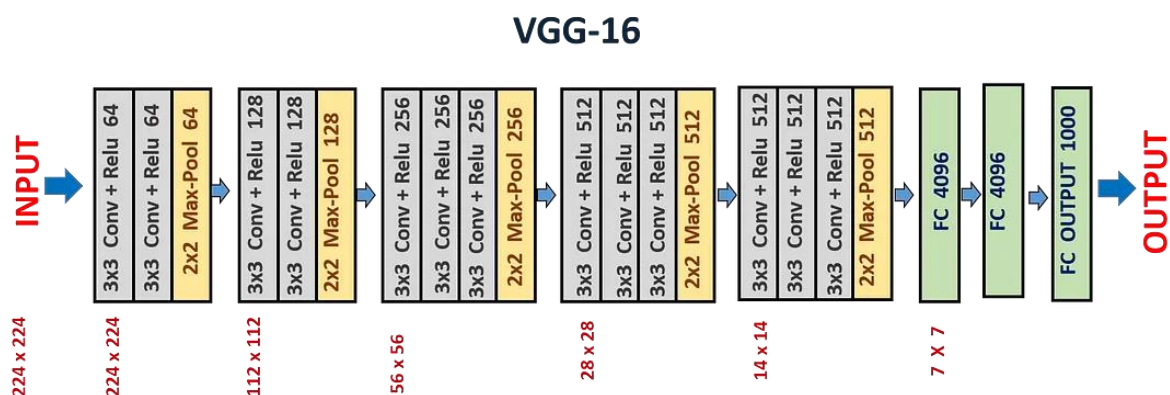
El modelo VGG16 está compuesto por 16 capas, distribuidas en 13 capas convolucionales y tres capas completamente conectadas (Fully Connected), todas ellas con filtros de tamaño 3×3 . Las capas convolucionales se agrupan en cinco bloques, cada uno de los cuales está seguido por una capa de agrupación máxima (MaxPooling) que reduce la dimensión espacial de las representaciones intermedias (Ali et al., 2021). VGG16 utiliza la función de activación ReLU para introducir no linealidad y mejorar el rendimiento del

modelo, mientras que la función Softmax se aplica en las capas finales para llevar a cabo la clasificación (Kim et al., 2021).

El primer bloque está conformado por la primera y segunda capa convolucional, cada una con 64 filtros, lo que transforma la imagen de entrada a una dimensión de $224 \times 224 \times 64$. La salida de estas capas es posteriormente enviada a una capa de agrupación máxima de tamaño 2×2 . El segundo bloque incluye la tercera y cuarta capa convolucional, ambas con 128 filtros, seguida también por una capa MaxPooling de 2×2 , reduciendo la salida a $56 \times 56 \times 128$. Las capas quinta, sexta y séptima constituyen el tercer bloque, empleando 256 filtros cada una y seguidas igualmente por una capa de agrupación máxima. Las capas comprendidas entre la octava y la decimotercera conforman dos bloques adicionales, cada uno con tres capas convolucionales que utilizan 512 filtros; en ambos casos, se aplica una capa MaxPooling tras cada grupo. Las capas decimocuarta y decimoquinta son completamente conectadas (Fully Connected), cada una con 4,096 unidades, y finalmente, la capa decimosexta corresponde a una capa de salida Softmax con 1,000 unidades (Philip et al., 2023). La arquitectura básica de VGG16 se presenta en la Figura 5.

Figura 5

Arquitectura Básica de VGG16



Nota. Extraído de The Architecture and Implementation of VGG-16 [Fotografía], por Vaibhav Khandelwal, 2020.

ResNet50. ResNet50 es una red neuronal convolucional de alto rendimiento compuesta por una arquitectura residual profunda de 50 capas, que mediante sus bloques de aprendizaje residual, ha permitido superar limitaciones presentes en arquitecturas más tradicionales, llevando la clasificación basada en imágenes a un nivel superior (Mayya et al., 2024).

Uno de los principales aportes de esta arquitectura es la resolución del problema del gradiente evanescente, el cual afecta al entrenamiento de redes profundas. Para mitigar este problema, ResNet50 incorpora conexiones residuales, que permiten combinar la salida de una capa convolucional con su entrada original, lo cual mejora la eficiencia del entrenamiento (Alkannad et al., 2025).

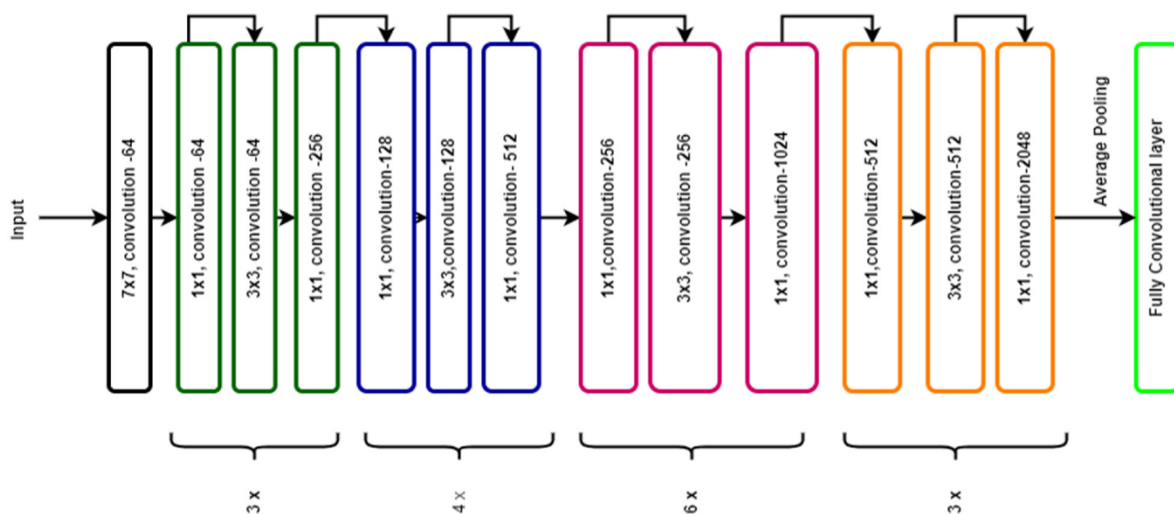
Los componentes fundamentales de ResNet50 son los bloques identidad y convolucional, los cuales se apoyan en funciones de activación ReLU y en técnicas de normalización por lotes (Batch Normalization). La salida final del modelo se obtiene mediante una función de activación Softmax (Kim et al., 2021).

La red inicia con una capa convolucional que emplea un kernel de 7×7 y 64 filtros, seguida por una capa de agrupación máxima (MaxPooling). Las siguientes capas convolucionales son de tamaño 1×1 , 3×3 y nuevamente 1×1 , con 64, 64 y 256 filtros respectivamente. Este bloque se repite tres veces, obteniendo nueve capas. A continuación, se incorpora un segundo conjunto de bloques con filtros de 1×1 , 3×3 y 1×1 , y 128, 128 y 512 filtros, respectivamente. Este bloque se repite cuatro veces, dando lugar a doce capas adicionales. El siguiente bloque introduce capas con 256, 256 y 1024 filtros, también en configuración de 1×1 , 3×3 y 1×1 , repetidas seis veces, generando dieciocho capas más. Finalmente, se añade un bloque con filtros de 512, 512 y 2048, en configuración de 1×1 , 3×3 y 1×1 , repetido tres veces, lo cual suma otras nueve capas a la arquitectura. La red finaliza

con una capa de agrupación promedio (Average Pooling), seguida de una capa completamente conectada con 1000 unidades. La salida final se obtiene mediante una función de activación Softmax (Philip et al., 2023). La arquitectura básica de ResNet50 se presenta en la Figura 6.

Figura 6

Arquitectura Básica de ResNet50



Nota. Extraído de A Comparative Study on Crack Detection in Concrete Walls Using Transfer Learning Techniques, por Philip et al., Journal of Composites Science, 7(4), 169, 2023.

Métricas de Evaluación del Modelo. Para evaluar el desempeño de los modelos ResNet50 y VGG16 se emplearon las métricas de exactitud (accuracy), precisión (precision) y sensibilidad (recall), calculadas durante las fases de entrenamiento y validación mediante las funciones integradas en Keras/TensorFlow. Estas métricas se derivan de la matriz de confusión, cuyos componentes clave son:

- **TP (Verdaderos Positivos):** Fisuras correctamente identificadas.
- **TN (Verdaderos Negativos):** Zonas sin fisuras correctamente clasificadas.

- **FP (Falsos Positivos):** Zonas sin fisuras erróneamente marcadas como fisuras.
- **FN (Falsos Negativos):** Fisuras reales no detectadas.

Las fórmulas aplicadas fueron las siguientes:

Exactitud (Accuracy). Mide el porcentaje total de predicciones correctas, tanto fisuras como zonas sanas, respecto al total de imágenes evaluadas.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precisión (Precision). Indica la proporción de verdaderas fisuras detectadas por el modelo entre todas las zonas marcadas como fisuras.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Sensibilidad (Recall). Representa la capacidad del modelo para identificar correctamente todas las fisuras presentes en las imágenes.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Parámetros de Entrenamiento. Se aplicaron configuraciones de entrenamiento y técnicas de regularización homogéneas en la medida de lo posible. El tamaño de entrada de las imágenes fue de 224×224 píxeles para ambos modelos, y se estableció un tamaño de lote (batch size) de 32. En el contexto del aprendizaje profundo, una época se define como un ciclo completo en el que el modelo ha procesado todas las muestras del conjunto de entrenamiento una vez. Cada época permite que el modelo ajuste sus pesos internos a partir de los errores cometidos, mejorando progresivamente su capacidad de generalización. El número total de épocas que se usó fue de 15, divididas en dos fases: 10 épocas iniciales para entrenar únicamente las capas nuevas añadidas a la base preentrenada y 5 épocas adicionales dedicadas al fine-tuning de capas específicas del modelo base. La función de pérdida

empleada fue `binary_crossentropy`, adecuada para tareas de clasificación binaria, y el optimizador Adam, con una tasa de aprendizaje de 0.0001 durante la primera fase y de 0.00001 durante el fine-tuning.

En ambos casos, se aplicaron estrategias para evitar el sobreajuste. Se implementó un esquema de aumento de datos (data augmentation) con transformaciones como rotaciones de 10 grados, desplazamientos horizontales y verticales del 10%, zoom del 15% y ajustes de brillo. Además, se incluyó una capa de Dropout (50%), regularización L2 (factor 0.01) en capas densas y Batch Normalization para mejorar la estabilidad del entrenamiento. Para optimizar el entrenamiento, en ambos modelos se emplearon los mismos mecanismos de control (callbacks): ModelCheckpoint, para conservar automáticamente la versión del modelo que alcanzó el mejor rendimiento sobre el conjunto de validación; EarlyStopping, para monitorear el progreso y finalizar el entrenamiento ante la ausencia de mejoras significativas durante 5 épocas consecutivas; y ReduceLROnPlateau, para ajustar la tasa de aprendizaje reduciéndola cuando deja de disminuir la pérdida de validación.

En cuanto a las características particulares de cada arquitectura, el modelo VGG16 utilizó una capa de Flatten para transformar la salida convolucional en un vector unidimensional antes de las capas densas, mientras que ResNet50 empleó una capa de GlobalAveragePooling2D, que permite reducir la dimensionalidad de forma más eficiente al promediar espacialmente las características. Durante la fase de fine-tuning, en VGG16 se congelaron las primeras 15 capas del modelo base, habilitando únicamente el entrenamiento de las capas superiores. En el caso de ResNet50, se descongeló aproximadamente la mitad superior del modelo preentrenado, permitiendo su actualización sobre los nuevos datos. Además, VGG16 incluyó capa intermedia de 1024 neuronas con regularización L2, seguida de otra de 512 neuronas antes de la capa de salida. Por su parte, ResNet50 implementó una única capa densa de 512 neuronas con regularización, previa a la capa final de clasificación.

Fase 4: Validación del Modelo (Aplicación en Campo)

Para el proceso de validación, se capturaron un total de 356 imágenes del Puente del Redondel San Juan, de las cuales 325 fueron obtenidas con una cámara Sony Alpha-6100 PZ 16-50mm F3.5-5.6 OSS a una resolución de 6024×4024 píxeles, y 31 imágenes adicionales fueron tomadas con un dron DJI Phantom 4 RTK a una resolución de 5472×3648 píxeles. Todas las imágenes fueron recortadas en subimágenes de 256×256 píxeles. A partir de este conjunto, se seleccionaron 2,000 imágenes para conformar la base de datos de validación, distribuidas en un 60 % con presencia de fisuras y un 40 % sin fisuras. Este conjunto fue utilizado para validar los modelos CNN en un entorno estructural real. En la Figura 7 se ilustra la captura de imágenes de las fisuras para su posterior análisis digital, y se evidencia el uso del fisurómetro para medir manualmente el ancho de fisuras.

Figura 7

Proceso de Levantamiento de Datos para Conjunto de Validación

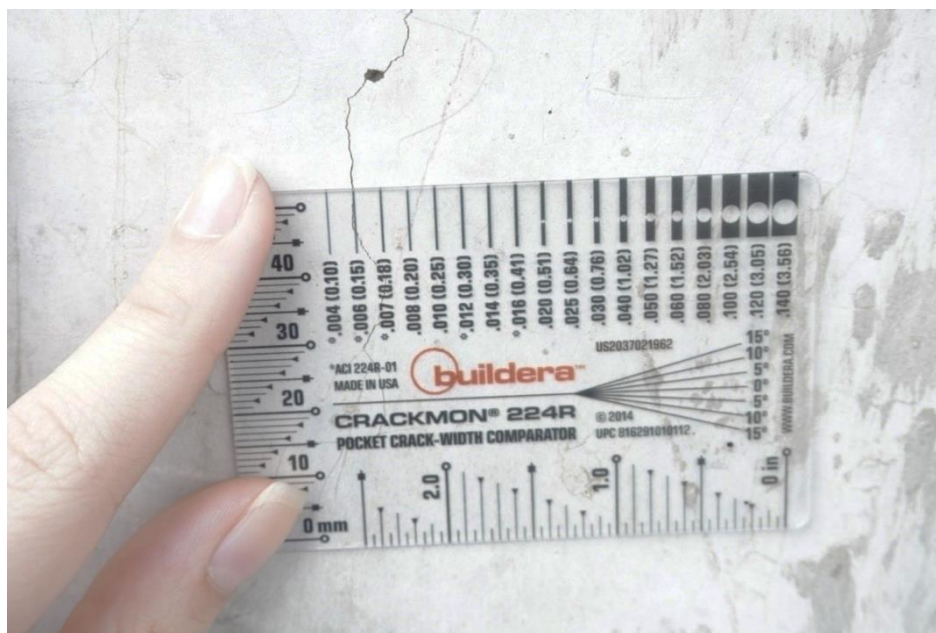


Nota. La imagen izquierda muestra el levantamiento fotográfico realizado para construir la base de datos de validación. La imagen derecha evidencia el uso de un fisurómetro para medir el ancho de fisuras en campo.

Durante las actividades de captura en campo para la validación, también se realizaron mediciones complementarias con fisurómetro en elementos estructurales del puente. Como se observa en la Figura 8, se identificó una fisura de 0.15 mm en un elemento principal, valor que coincide exactamente con el límite máximo permitido por la normativa ACI 224R-01 ACI, (2001) para estructuras expuestas a ambientes marinos o ciclos de humedecimiento-secado, lo que indica que, aun cuando la fisura no parezca crítica visualmente, su magnitud puede implicar riesgos importantes a mediano plazo, especialmente en una estructura que se encuentra inconclusa y sin enlucido, como es el caso del puente.

Figura 8

Medición de fisura en Elemento Estructural del Puente del Redondel San Juan



Nota. La imagen muestra una fisura de 0.15 mm detectada en un componente principal del puente, medida con un fisurómetro durante las inspecciones en campo.

Materiales

Para la captura de imágenes, se empleó una cámara Sony Alpha-6100 PZ 16-50mm F3.5-5.6 OSS, la cual cuenta con una distancia focal de 16-50 mm (equivalente en 35 mm: 24-75 mm), sensor CMOS APS-C de 24.2 megapíxeles, un sistema enfoque automático híbrido con 425 puntos de detección de fase y contraste, y estabilización óptica de imagen (OSS) incorporada que minimiza el desenfoque (SONY ELECTRONICS INC., s.f.-b). El cuerpo del equipo (modelo ILCE-6100) corresponde a la versión 01, mientras que el objetivo utilizado pertenece a la versión 02.

Figura 9

Cámara Sony Alpha-6100



Nota. Extraído de Alpha 6100 APS-C Mirrorless Interchangeable-Lens Camera + 16-50mm Lens [Fotografía], por SONY ELECTRONICS INC., s.f.

Para el procesamiento de imágenes y el entrenamiento de los modelos de redes neuronales convolucionales, se utilizó el entorno de desarrollo PyCharm, versión 2025.1.1.1, junto con la biblioteca TensorFlow/Keras.

Figura 10

Entorno de Desarrollo PyCharm 2025.1.1.1



Nota. Extraído de Logo of PyCharm [Fotografía], por JetBrains, 2020.

Las tareas de procesamiento y entrenamiento fueron ejecutadas en una computadora portátil marca Lenovo IdeaPad 3, equipada con un procesador AMD Ryzen 7 5700U, gráficos integrados Radeon, 12 GB de memoria RAM y sistema operativo Windows 11 de 64 bits.

Figura 11

Computadora Portátil Lenovo IdeaPad 3



Nota. Extraído de Lenovo IdeaPad 3 Notebook 15,6" [Fotografía], de Lenovo, s.f.

Para la captura de imágenes aéreas del Puente del Redondel San Juan, se utilizó un dron DJI Phantom 4 RTK, mostrado en la Figura 12.

Figura 12

Dron DJI Phantom 4 RTK



Nota. Extraído de DJI Phantom 4 RTK [Fotografía], por DJI, s.f.

Población y Muestra

Descripción de la Obra Seleccionada

Para la validación práctica del modelo de red neuronal convolucional desarrollado, se seleccionó como caso de estudio el Puente Redondel San Juan, estructura de hormigón armado ubicada en una vía principal de alto tráfico vehicular, la cual se puede apreciar en la Figura 13.

Figura 13*Puente Redondel San Juan*

Nota. Imagen del Puente Redondel San Juan capturada con un Dron DJI Phantom 4 RTK.

La elección de esta obra se justifica en sus características particulares, las cuales la convierten en un escenario idóneo para la aplicación del sistema de detección automatizada de fisuras. Al tratarse de una estructura inconclusa y abandonada, presenta superficies de hormigón vistas sin revestimientos, lo que permite la observación directa de fisuras en el hormigón.

Desde la perspectiva técnica, el puente presenta un alto valor como caso de estudio debido a su condición estructural actual. La exposición prolongada a factores ambientales sin mantenimiento ha generado un patrón de deterioro característico en estructuras de hormigón que incluye diversos tipos de fisuras, lo cual permite evaluar la capacidad del modelo para identificar fisuras en el hormigón bajo condiciones reales.

La selección de esta obra también consideró aspectos operativos y de impacto social. Como infraestructura ubicada en una zona de alta circulación vehicular, su evaluación representa un caso con relevancia práctica inmediata. La accesibilidad del puente facilita la

toma sistemática de imágenes en diferentes condiciones de iluminación y perspectivas, permitiendo una validación exhaustiva del modelo.

Número de las Imágenes

Dado que no existe una normativa establecida por un ente regulador para definir la cantidad de imágenes requeridas, se llevó a cabo una revisión bibliográfica del estado del arte. Los hallazgos derivados de este análisis se presentan de manera resumida en la Tabla 2.

Tabla 2

Análisis Bibliográfico de Cantidad de Imágenes y Rendimiento de Modelos

Autor	Imágenes	Resultados
B. Chen et al. (2023)	3,000	99.79%
Hui et al. (2025)	40,000	99%
Yamane & Chun (2020)	100	99%
Billah et al. (2019)	6,000	98.50%
Ma et al. (2021)	1,595	98.22%
Cha et al. (2017)	40,000	98%
Zoubir et al. (2022)	6,900	97.38%
Choi, Bae, et al. (2024)	1,203	96,50%

Nota. La tabla resume los hallazgos de la revisión bibliográfica sobre el estado del arte en el campo de la detección de fisuras mediante CNN, destacando la cantidad de imágenes utilizadas y los resultados obtenidos por diferentes autores.

De las investigaciones revisadas, se seleccionaron aquellas que reportaron los mejores resultados de precisión en la detección de fisuras mediante redes neuronales convolucionales. Aunque algunos estudios utilizaron bases de datos significativamente extensas, de hasta 40,000 imágenes (Cha et al. 2017; Hui et al. 2025), estos no fueron considerados en el cálculo promedio debido a limitaciones logísticas inherentes al presente trabajo de titulación, como restricciones de tiempo, recursos técnicos y condiciones de seguridad para la recolección de

datos. A partir de los estudios restantes que mantuvieron una alta precisión con conjuntos de datos más moderados, se determinó que un tamaño de base de datos de aproximadamente 8,000 imágenes resulta adecuado para alcanzar un equilibrio entre calidad, rendimiento del modelo y viabilidad operativa. Esta cantidad se sitúa en un punto intermedio respecto a las investigaciones consultadas, evitando tanto el uso de muestras extremadamente reducidas como el manejo de volúmenes masivos que superan las capacidades del presente proyecto.

Resultados

Levantamiento de Datos

La Tabla 3 resume el proceso de captura y procesamiento de imágenes empleado para la creación de la base de datos utilizada en el entrenamiento y validación de los modelos CNN. Se presenta el número de imágenes originales capturadas en campo, sus resoluciones iniciales y la cantidad de imágenes seleccionadas tras el proceso de recorte. Las imágenes finales, utilizadas como entradas para el modelo, fueron procesadas y estandarizadas con una resolución de 256×256 píxeles.

Tabla 3

Resultados de la Captura y Procesamiento de Imágenes para la Base de Datos CNN

Fase	Imágenes capturadas	Resolución original	Imágenes seleccionadas	Resolución final
Entrenamiento	715	6024×4024 px	6,000	256×256 px
Validación	356 (325 cámara + 31 dron)	6024×4024 px / 5472×3648 px	2,000	256×256 px
Total	1,071	—	8,000	—

Nota. La tabla detalla las fases de captura y procesamiento de imágenes para la creación de la base de datos utilizada en el entrenamiento y validación de los modelos de CNN.

La fase de entrenamiento se basó en un conjunto amplio de imágenes capturadas desde distintas estructuras de hormigón, mientras que la fase de validación se concentró exclusivamente en un caso real: el Puente del Redondel San Juan. Esta estrategia permitió construir un conjunto de entrenamiento variado y representativo, y, al mismo tiempo, evaluar el modelo en condiciones reales. Las imágenes capturadas fueron procesadas digitalmente mediante recortes sistemáticos para estandarizarlas a una resolución de 256×256 píxeles, adecuada para su uso como entrada en modelos de redes neuronales convolucionales. Esta

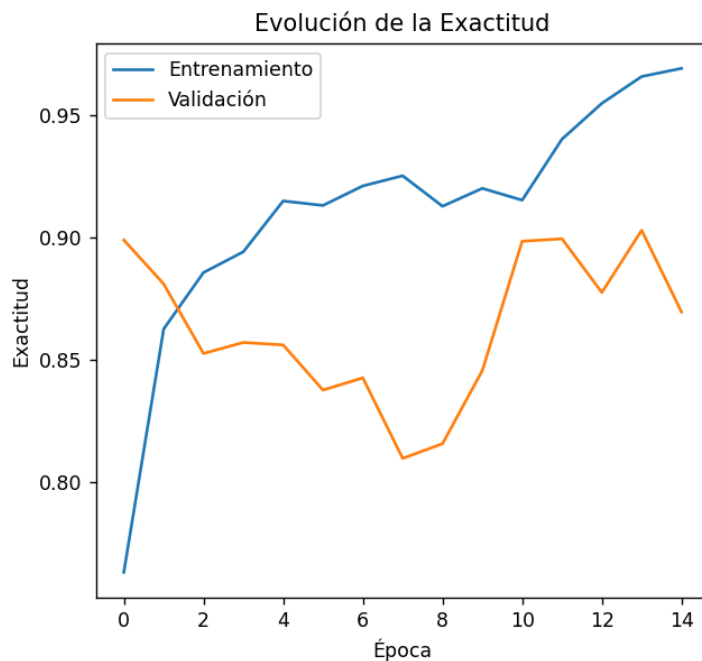
resolución equilibró el nivel de detalle visual requerido para detectar fisuras y la eficiencia computacional necesaria para el entrenamiento.

Entrenamiento de CNN

ResNet50

Durante el proceso de entrenamiento del modelo ResNet50, se registraron y analizaron las principales métricas de desempeño a lo largo de 15 épocas. Los gráficos obtenidos permiten observar el comportamiento del modelo respecto a la exactitud, la precisión y la sensibilidad, además de la función de pérdida, en ambos conjuntos de entrenamiento y de validación.

En la Figura 14, se representa la evolución de la exactitud (accuracy). La curva azul corresponde al conjunto de entrenamiento y muestra una mejora continua, alcanzando valores superiores al 90 % a partir de la época 4 y manteniéndose estable en las últimas iteraciones. Por otro lado, la curva naranja, correspondiente al conjunto de validación, presenta una tendencia menos estable: tras una ligera mejora inicial, desciende entre las épocas 2 y 8. Sin embargo, a partir de la época 9, se evidencia una recuperación progresiva, alcanzando valores cercanos al 90 %, lo cual sugiere una mejora en la capacidad de generalización gracias al ajuste fino aplicado en la segunda etapa del entrenamiento. Las fluctuaciones en la curva indican que aún existen variaciones sensibles en el rendimiento del modelo cuando se enfrenta a datos no vistos.

Figura 14*Evolución de la Exactitud de ResNet50*

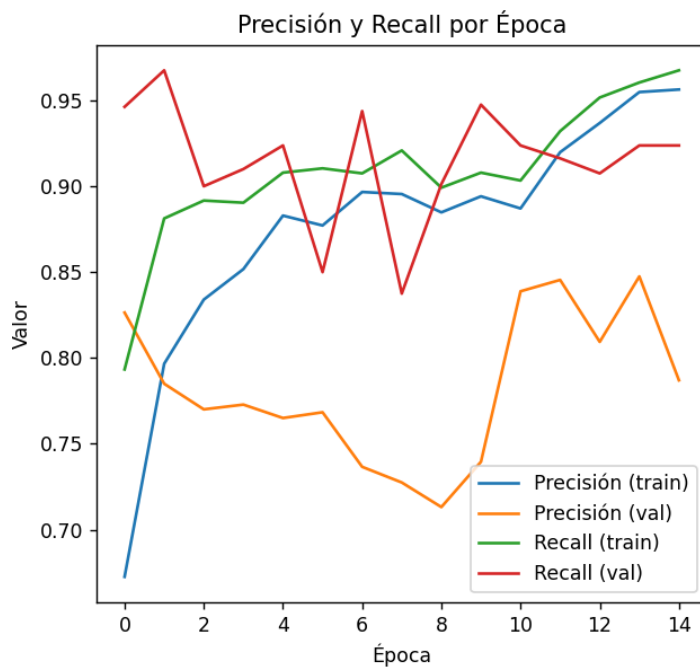
Nota. Figura generada en Pycharm de la evolución de la exactitud del modelo ResNet50 durante el entrenamiento y validación a lo largo de 15 épocas.

La Figura 15 muestra la evolución conjunta de la precisión (precision) y la sensibilidad (recall). En el conjunto de entrenamiento, se ve un crecimiento constante hasta superar el umbral del 90%, estabilizándose en ese rango, lo que refleja que el modelo ha aprendido a diferenciar correctamente entre imágenes fisuradas y no fisuradas.

En el conjunto de validación se observan mayores oscilaciones. descendió progresivamente entre las épocas 2 y 9, indicando un aumento en los falsos positivos durante esa fase. A partir de la época 10, se produjo una recuperación que alcanzó valores cercanos al 87 %. La sensibilidad presentó picos irregulares durante las primeras épocas, y una ligera tendencia decreciente hasta la época 8, estabilizándose posteriormente entre el 85 % y 95 %.

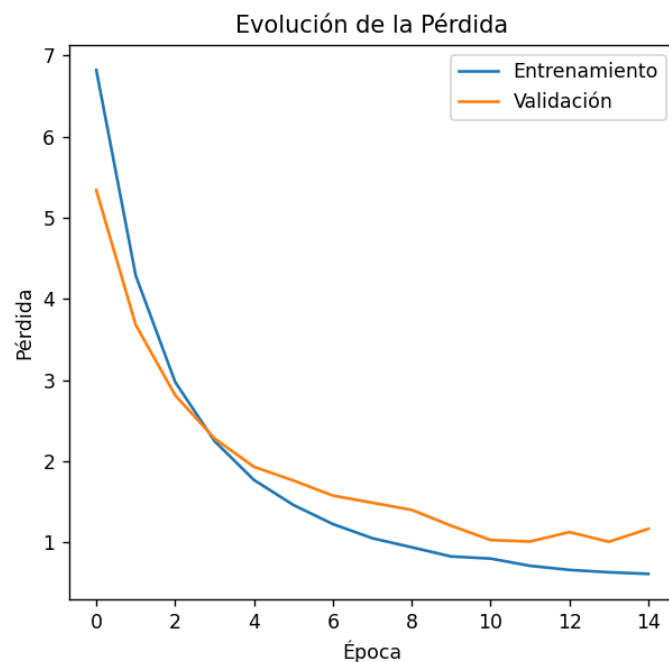
Figura 15

Evolución de la Precisión y Sensibilidad de ResNet50



Nota. Figura generada en Pycharm de la evolución de la precisión y sensibilidad del modelo ResNet50 durante el entrenamiento y validación a lo largo de 15 épocas.

En la Figura 16, se analiza la evolución de la función de pérdida. Se observa una reducción rápida y consistente en ambas curvas durante las 5 primeras épocas. Desde la época 6 en adelante, esta métrica continúa disminuyendo de manera estable en el conjunto de entrenamiento, mientras que en la validación se reduce de forma más moderada y se estabiliza a partir de la época 10.

Figura 16*Evolución de la Pérdida de ResNet50*

Nota. Figura generada en Pycharm de la evolución de la función de pérdida del modelo ResNet50 durante el entrenamiento y validación a lo largo de 15 épocas.

Al concluir el entrenamiento, el modelo ResNet50 alcanzó los siguientes resultados sobre el conjunto de validación: una exactitud (accuracy) del 90.80 %, una precisión (precision) del 86.93 % y una sensibilidad (recall) del 90.62 %. Estos valores reflejan un desempeño sólido, con un buen equilibrio entre identificar los casos positivos correctamente, lo que se refleja en un alto valor de sensibilidad, y minimizar errores en las predicciones, representado por la precisión.

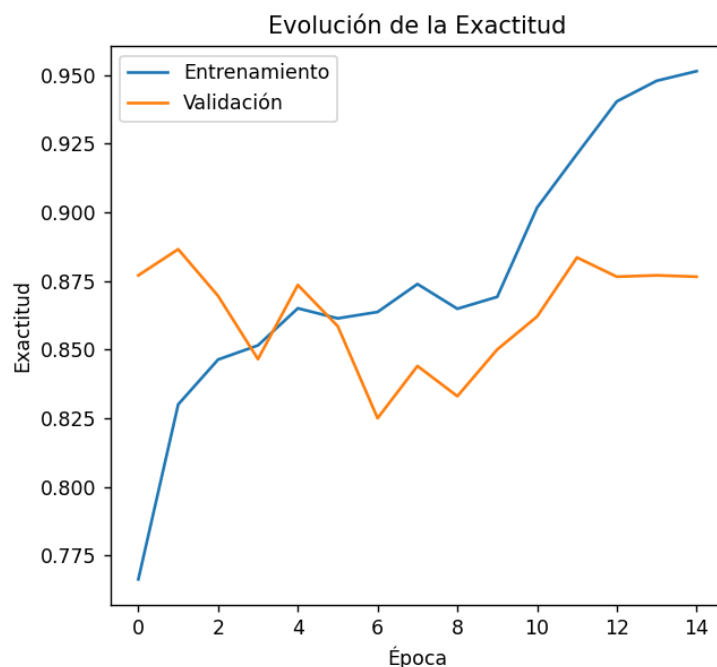
VGG16

Durante el proceso de entrenamiento del modelo VGG16, también se registraron y analizaron las métricas de exactitud, la precisión y la sensibilidad a lo largo de 15 épocas. Los resultados se presentan a continuación a través del análisis de tres gráficos principales.

En la Figura 17, se muestra la evolución de la exactitud de VGG16. Desde las primeras épocas, se observa un aumento progresivo en la exactitud del entrenamiento, que culmina en un valor cercano al 95 %. La exactitud en validación experimenta una ligera oscilación entre las épocas 3 y 8, con un descenso temporal hasta un mínimo en la época 6. A partir de ese punto, se recupera de manera gradual, alcanzando una estabilización alrededor del 88 %.

Figura 17

Evolución de la Exactitud de VGG16



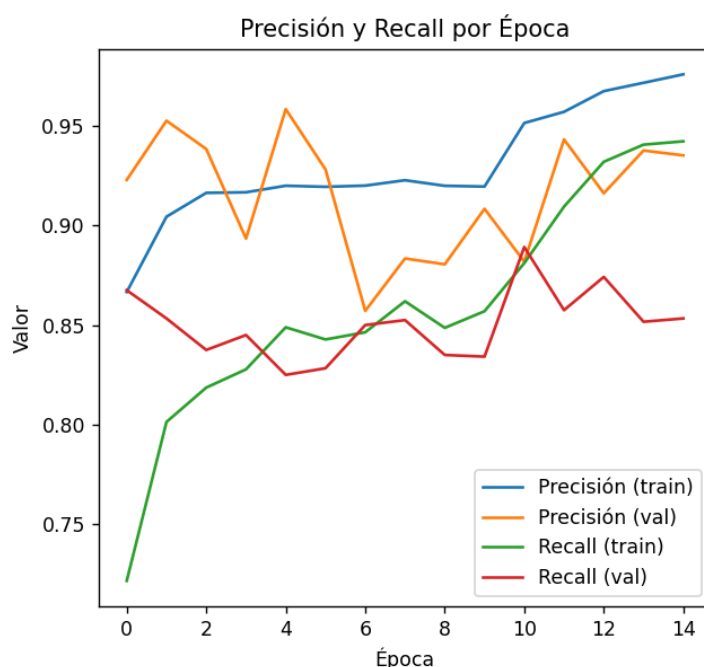
Nota. Figura generada en Pycharm de la evolución de la exactitud del modelo VGG16 durante el entrenamiento y validación a lo largo de 15 épocas.

La Figura 18 representa la evolución de las métricas de precisión y sensibilidad (recall) en función de las épocas. En el conjunto de entrenamiento, ambas métricas mejoran consistentemente con el avance del entrenamiento, especialmente la precisión, que alcanza valores cercanos al 97 %.

Por el contrario, las curvas correspondientes a la validación presentan una variabilidad significativa. La precisión alcanza valores altos desde las primeras épocas, aunque presenta oscilaciones entre el 87 % y el 96 %, reflejando una sensibilidad a la presencia de falsos positivos en distintas fases del entrenamiento. En cuanto a la sensibilidad en validación, se observa una leve tendencia decreciente durante las primeras épocas, seguida de una estabilización entre el 85 % y el 88 %.

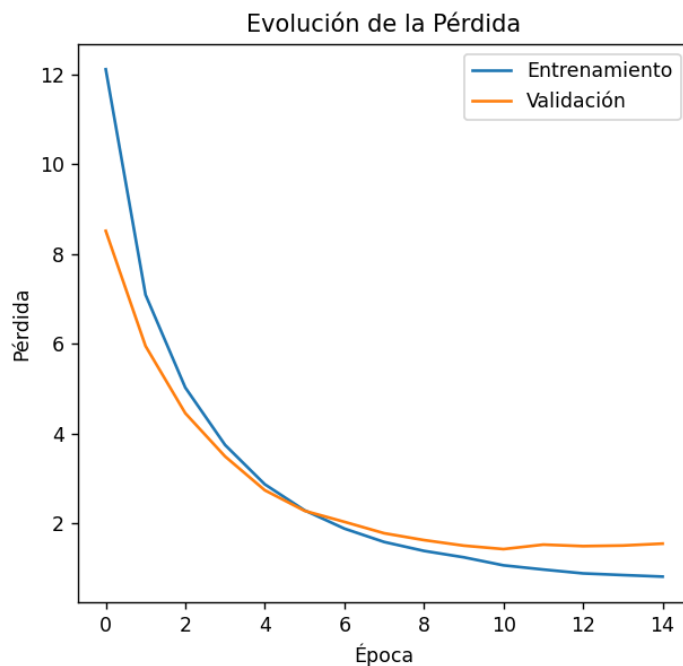
Figura 18

Evolución de la Precisión y Sensibilidad de VGG16



Nota. Figura generada en Pycharm de la evolución de la precisión y sensibilidad del modelo VGG16 durante el entrenamiento y validación a lo largo de 15 épocas.

La Figura 19 muestra la evolución de la función de pérdida. Ambas curvas, tanto la de entrenamiento como la de validación, presentan un descenso constante durante las primeras 10 épocas. La pérdida en entrenamiento disminuye de forma más marcada y continua, mientras que en validación se estabiliza entre las épocas 10 y 15, con una ligera tendencia al alza en las últimas iteraciones.

Figura 19*Evolución de la Pérdida de VGG16*

Nota. Figura generada en Pycharm de la evolución de la función de pérdida del modelo ResNet50 durante el entrenamiento y validación a lo largo de 15 épocas.

Al finalizar el proceso de entrenamiento, el modelo VGG16 alcanzó los siguientes resultados de desempeño sobre el conjunto de validación: una exactitud del 88.95 %, una precisión del 96.75 % y una sensibilidad del 84.42 %. Estos valores reflejan un modelo con alta capacidad para evitar falsos positivos, evidenciado por su precisión, aunque con un rendimiento ligeramente inferior respecto a su capacidad para detectar todos los casos positivos (fisuras), lo que se evidencia en su sensibilidad.

Comparación de Resultados de CNN

A continuación, se presenta en la Tabla 4 un resumen comparativo de los resultados obtenidos por VGG16 y ResNet50 en el conjunto de validación.

Tabla 4

Resultados Comparativos de los Modelos VGG16 y ResNet50

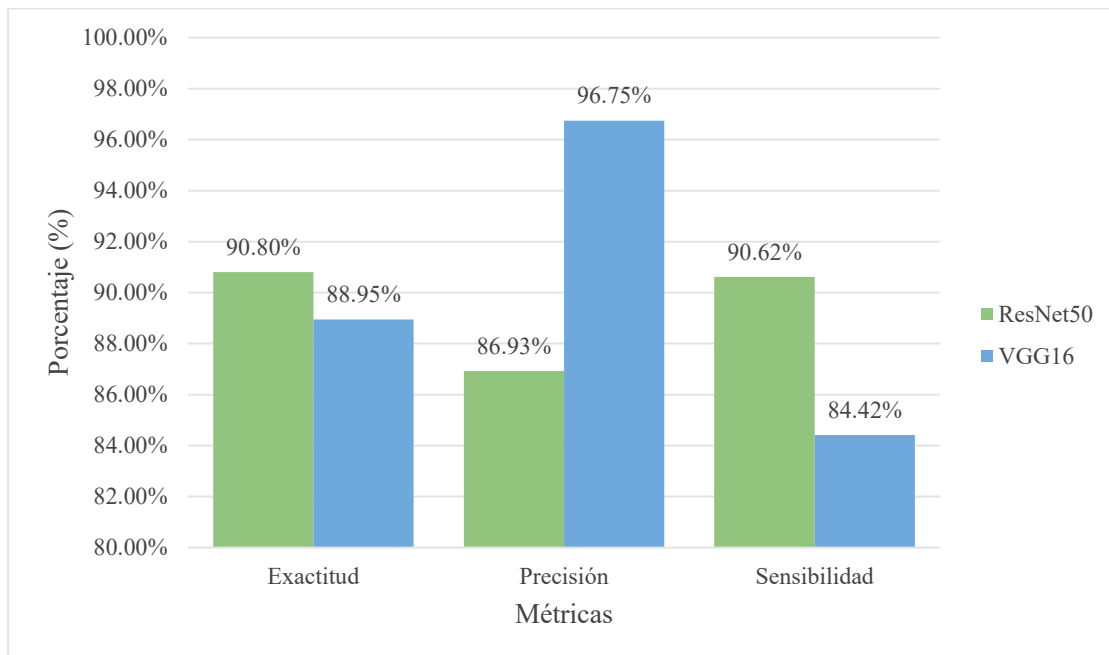
Métricas	ResNet50	VGG16	Interpretación
Exactitud	90.80%	88.95%	ResNet50 es ligeramente mejor en aciertos globales.
Precisión	86.93%	96.75%	VGG16 tiene menos falsas alarmas.
Sensibilidad	90.62%	84.42%	ResNet50 omite menos fisuras reales.

Nota. La tabla presenta la evaluación comparativa de los modelos ResNet50 y VGG16. Las métricas de Exactitud, Precisión y Sensibilidad se muestran para cada modelo, junto con una interpretación de los resultados obtenidos.

La Figura 20 presenta una comparación gráfica del desempeño de los modelos de CNN VGG16 y ResNet50. Como se puede observar, el modelo ResNet50 alcanzó una mayor exactitud (90,80 %) y sensibilidad (90,62 %) en comparación con VGG16, lo que indica una mayor capacidad para identificar correctamente imágenes con y sin fisuras, así como para detectar la mayoría de las fisuras presentes. En cambio, VGG16 superó a ResNet50 en precisión, con un valor de 96,75 %, lo cual refleja una menor cantidad de falsos positivos.

Figura 20

Comparación del Desempeño entre VGG16 y ResNet50 en el Conjunto De Validación



Nota. Comparación visual de los resultados de experimentación de CNN para los modelos VGG16 y ResNet50 en el conjunto de validación, evaluando las métricas de Exactitud, Precisión y Sensibilidad. Los valores porcentuales indican el rendimiento de cada modelo en cada métrica.

Discusión

Uno de los principales desafíos que enfrenta la industria de la construcción no es únicamente técnico, sino también cultural, ya que existe una resistencia al cambio y a la adopción de nuevas tecnologías. Esto motivó el enfoque de este trabajo de titulación, que busca no solo proponer una solución técnica al problema de la detección de fisuras en estructuras de hormigón, sino también contribuir al proceso de modernización del sector de la arquitectura, ingeniería y construcción (AEC). Al comparar el lento avance tecnológico de esta industria con el progreso acelerado de otros sectores, resulta evidente la necesidad de impulsar iniciativas que integren la inteligencia artificial, la visión por computadora y la automatización en los procesos constructivos y de mantenimiento.

Este estudio representa una iniciativa para introducir herramientas innovadoras dentro del sector de AEC, una industria que históricamente ha mostrado una baja tasa de adopción tecnológica en comparación con sectores como la manufactura avanzada o el sector automotriz, manteniéndose en gran medida anclada a métodos tradicionales (McKinsey & Company, 2017). La falta de innovación se refleja en la baja productividad del sector, la cual ha sido señalada como un desafío global con impactos significativos en su desempeño general (Infrastructure and Projects Authority, 2017).

En este contexto, la automatización de la inspección de fisuras surge como una necesidad creciente. Los métodos tradicionales para este proceso se basan principalmente en inspecciones visuales manuales que, aunque han sido ampliamente utilizados durante décadas, presenta limitaciones significativas, ya que demandan mayor cantidad de tiempo, recursos y mano de obra, que además, son susceptibles a errores humanos (Hui et al., 2025). En contraste, los métodos basados en redes neuronales convolucionales han demostrado ser altamente eficaces en la detección de fisuras, incluso en tareas de clasificación complejas, y gracias al aprendizaje por transferencia, estas redes pueden adaptarse fácilmente a distintos

conjuntos de datos, lo que mejora su versatilidad y aplicabilidad en distintos contextos de análisis de fisuras (Fu et al., 2020).

El enfoque adoptado en este trabajo de titulación se alinea con los esfuerzos globales por modernizar la industria AEC, no solo buscando mejorar la eficiencia del diagnóstico estructural, sino también fomentando la integración de nuevas tecnologías en una industria que, según el informe de World Economic Forum (2016), se encuentra rezagada en términos de digitalización.

Comparación Metodológica con Investigaciones Previas

La construcción de una base de datos propia a partir de imágenes capturadas en condiciones reales representa uno de los principales aportes metodológicos de esta investigación. A diferencia de estudios que se basan en conjuntos de datos públicos, muchas veces homogéneos y generados en condiciones controladas, el uso de imágenes provenientes de escenarios reales introduce una mayor variabilidad en términos de textura, iluminación, ángulo de captura y presencia de elementos no estructurales, como marcas de acabado o pisadas en el hormigón. Esta diversidad en las condiciones de captura ofrece un entorno más representativo de los desafíos que se presentan en la práctica profesional de la inspección estructural.

La decisión de utilizar una cámara profesional y un dron permitió abarcar diferentes escalas y tipos de superficie, favoreciendo la obtención de una base de datos robusta. La posterior segmentación en imágenes de 256×256 píxeles, no solo estandarizó el formato para el entrenamiento de los modelos, sino que también permitió una mayor densidad de muestras, incrementando la eficacia del aprendizaje automático sin comprometer la calidad de la información visual.

Si bien el tamaño total del conjunto de datos de 8,000 imágenes se encuentra en un rango medio en comparación con investigaciones similares, su valor reside en su autenticidad y complejidad. Este enfoque aporta un marco más desafiante para el entrenamiento de modelos de redes neuronales convolucionales, lo cual pone a prueba su capacidad de generalización en contextos reales.

Este enfoque ha sido adoptado en otros estudios recientes que, al igual que esta investigación, optaron por construir sus propias bases de datos. Zhu & Song (2020) recolectaron 40,610 imágenes utilizando cámaras de escaneo lineal con sensores CCD, clasificando las imágenes entre normales, fisuradas y con otros tipos de defectos. Ojeda et al. (2023) construyeron una base de datos de 2,650 imágenes a partir de ensayos en cilindros de hormigón en laboratorio, mientras que Choi, Park, et al. (2024) utilizaron 1,500 imágenes térmicas para el análisis de fisuras. Por su parte, L. Chen et al. (2023) utilizaron un conjunto de 6,600 imágenes, Zoubir et al. (2021) 6,938, y Nugraheni et al. (2023) y Yao et al. (2021) trabajaron con 10,000 imágenes cada uno.

En contraste, diversos estudios han empleado bases de datos públicas ampliamente conocidas. Por ejemplo, Junior et al. (2024), Hui et al. (2025), Kim et al. (2021) y Zadeh et al. (2023) utilizaron el conjunto de datos de Özgenel & Gönenç Sorguç (2019), compuesto por 40,000 imágenes tomadas en edificios del campus de la Middle East Technical University (METU). Así mismo, investigaciones como las de Mayya et al. (2024) y Philip et al. (2023) utilizaron la base SDNET2018, desarrollada por Dorafshan et al. (2018), la cual contiene 56,092 imágenes, de las cuales 11,595 son de superficies sin fisuras y 2,025 con fisuras. Algunos estudios, como los de Ali et al. (2021) y Alkannad et al. (2025), combinaron ambas bases de datos en sus experimentos. Si bien estas bases son ampliamente utilizadas y permiten una rápida implementación de modelos, sus imágenes fueron generadas en

condiciones controladas, lo que limita su representatividad frente a condiciones reales de inspección.

Además, otros autores han utilizado diferentes fuentes de datos con técnicas de aumento de datos para expandir su base. Guo et al. (2020), por ejemplo, emplearon los conjuntos de Zhang et al. (2016) y Li et al. (2019), que contienen 40,000 y 2,068 imágenes, respectivamente. A través del uso intensivo de técnicas de data augmentation, incrementaron su base hasta alcanzar 130,000 y 6,069 imágenes, respectivamente. Este enfoque permite mejorar el aprendizaje del modelo, aunque introduce redundancia artificial que no necesariamente representa nuevas condiciones reales.

En lo que respecta a la organización del conjunto de datos, una de las decisiones metodológicas clave en las investigaciones basadas en aprendizaje automático es la forma en que se divide la base de datos entre entrenamiento, validación y prueba. Esta división influye directamente en la capacidad del modelo para generalizar su aprendizaje y evaluar su rendimiento de forma precisa y objetiva.

En esta investigación, se optó por una división en dos conjuntos: 6,000 imágenes para entrenamiento y 2,000 imágenes para validación, lo cual corresponde a una proporción del 75 % para entrenamiento y 25 % para validación. A diferencia de otros estudios, no se utilizó un conjunto separado de prueba, dado que el conjunto de validación fue generado a partir de imágenes exclusivas tomadas de una sola estructura real, el Puente del Redondel San Juan, lo cual permitió una validación práctica más exigente, enfocada en evaluar el modelo en un entorno real.

De forma similar, también existen investigaciones que, al igual que este trabajo de titulación, dividieron sus datos únicamente entre entrenamiento y validación o prueba. Mayya et al. (2024), Zhu & Song (2020), Guo et al. (2020), Alkannad et al. (2025) y Philip et al.

(2023) aplicaron divisiones en proporciones que oscilan entre 70/30 y 90/10, priorizando el entrenamiento y empleando una sola partición para validar o probar el rendimiento. En estos casos, la ausencia de un tercer conjunto suele responder a limitaciones prácticas o al uso de validación cruzada. Hui et al. (2025), por ejemplo, dividió su conjunto en 80 % de entrenamiento y 20 % de validación, sin realizar una evaluación separada posterior. Lo mismo se observó en Nugraheni et al. (2023), quienes aplicaron un esquema 70/30.

En comparación, varios trabajos en la literatura dividieron sus datos en tres subconjuntos: entrenamiento, validación y prueba. Por ejemplo, Junior et al. (2024) utilizó un esquema de 65 %-15 %-20 %, equivalente a 27,000 imágenes para entrenamiento, 6,000 para validación y 7,000 para prueba. Kim et al. (2021) también empleó esta estructura, asignando 32,000 imágenes al entrenamiento, y 4,000 tanto para validación como para prueba. De forma similar, (Zoubir et al., 2021, 2022) y Yao et al. (2021) implementaron divisiones de 70 %-10 %-20 %, reservando una parte específica del conjunto para pruebas finales independientes, lo cual permite una evaluación más robusta del modelo entrenado en múltiples fases.

Si bien el uso de tres conjuntos puede ofrecer una evaluación más detallada, la metodología de esta investigación compensa esa ausencia mediante una validación aplicada directamente a una estructura real, capturada en condiciones no controladas, lo que proporciona un enfoque alternativo con mayor representatividad práctica. En este sentido, el diseño experimental aquí propuesto permite evaluar el rendimiento de los modelos bajo condiciones operativas reales, lo cual otorga mayor relevancia al análisis de desempeño, aún sin una partición de prueba formal.

Comparación del Desempeño de Métricas de Modelos

Los resultados alcanzados en esta investigación reflejan la capacidad real de los modelos CNN implementados para identificar fisuras en condiciones prácticas. La

arquitectura VGG16 demostró un desempeño sólido en la tarea de detección de fisuras, alcanzando una exactitud del 88.95 %, lo que indica que, en términos generales, el modelo clasificó correctamente la mayoría de las imágenes evaluadas. Su precisión del 96.75 % evidenció una destacada capacidad para evitar falsos positivos, es decir, para identificar correctamente las imágenes que no contenían fisuras sin confundirlas con casos defectuosos. Sin embargo, su sensibilidad del 84.42 % fue relativamente menor, lo que sugiere que el modelo tuvo más dificultades para detectar todas las imágenes que sí presentaban fisuras, omitiendo una fracción de los casos verdaderamente positivos.

Por otra parte, el modelo ResNet50 mostró un desempeño más equilibrado entre sus métricas, con una exactitud del 90.80 % que refleja un rendimiento general sólido, indicando que clasificó correctamente una proporción alta de las imágenes evaluadas. La sensibilidad alcanzó un 90.62 %, lo que representa una alta capacidad del modelo para identificar correctamente las imágenes que en realidad contenían fisuras. Su precisión del 86.93 %, aunque ligeramente inferior a la de VGG16, sugiere que el modelo cometió más falsos positivos, es decir, clasificó algunas imágenes sin fisuras como si presentaran fisuras. Sin embargo, ResNet50 ofreció un mejor rendimiento general que VGG16 dentro de las condiciones reales de validación aplicadas en este estudio.

En conjunto, los resultados alcanzados cumplen el objetivo de evaluar el rendimiento de modelos CNN con datos reales, y demuestran que, a pesar de las condiciones menos controladas, los modelos pueden lograr niveles de precisión y sensibilidad competitivos.

Estudios como el de Junior et al. (2024) emplearon modelos preentrenados VGG16, VGG19 y ResNet50, logrando resultados notoriamente altos: exactitud de 99.77 %, precisión de 99.57% y sensibilidad de 99.90 % para VGG16, y exactitud de 99.27 % con precisión de 99.83 % y sensibilidad de 98.47% para ResNet50. Aunque superiores a los resultados de la

presente investigación, tales cifras fueron alcanzadas bajo condiciones experimentales muy controladas, con una base de datos de 40,000 imágenes extraídas de un único entorno universitario, lo que limita la variabilidad contextual. Asimismo, Zadeh et al. (2023) reportaron para ResNet50 una exactitud del 99.4 %, precisión de 98.8 % y sensibilidad de 99.9 %, utilizando técnicas de fine-tuning sobre una base de datos pública estandarizada. En contraste, los resultados de este trabajo de titulación fueron obtenidos con una base de datos de 8,000 imágenes reales, tomadas con cámara profesional y dron en distintas estructuras de hormigón, bajo condiciones variables de iluminación y textura, lo que añade complejidad al proceso de clasificación. La diferencia en la calidad y homogeneidad de los datos puede explicar parcialmente la disparidad en los resultados.

En Mayya et al. (2024), VGG16 alcanzó 92 % tanto en precisión como en sensibilidad mientras que ResNet50 obtuvo 86 % de precisión y 85 % de sensibilidad, cifras que se aproximan a las obtenidas en este trabajo de titulación que alcanzó resultados con VGG16 de 96.75 % de precisión y 84.42 % de sensibilidad; y con ResNet50 de 86.93 % y 90.62 %, respectivamente. Esta similitud metodológica, en el uso de modelos preentrenados y datos aumentados, permite considerar que el rendimiento alcanzado aquí es comparable a otras investigaciones recientes, incluso cuando estas utilizaron conjuntos de datos más amplios o mejor balanceados.

En el estudio de Ojeda et al. (2023), que trabajó también con imágenes recolectadas en laboratorio y clasificadas manualmente, se reportó una exactitud general del 91 % para ResNet50 y 90 % para VGG16, valores prácticamente equivalentes a los de esta investigación, lo que reafirma la efectividad de estos modelos en condiciones no ideales.

Por su parte, investigaciones como las de Guo et al. (2020) y Zhu & Song (2020), aunque implementaron CNN con arquitecturas personalizadas o VGG16 mejorado,

obtuvieron exactitudes de entre 96.91 % y 95.56%, y 83.76 %, respectivamente. Incluso en diseños a medida, el rendimiento puede ser similar o inferior al de modelos preentrenados ajustados adecuadamente, como los utilizados en la presente investigación.

Otras investigaciones han profundizado en la comparación entre distintas arquitecturas de redes neuronales convolucionales, incluyendo versiones personalizadas o modelos clásicos ampliamente difundidos en tareas de visión por computadora. Por ejemplo, Kim et al. (2021) desarrollaron una versión modificada del modelo LeNet-5 denominada OLeNet, la cual fue evaluada junto a VGG16, Inception y ResNet en tareas de detección de fisuras. Los resultados fueron sobresalientes en todos los casos, alcanzando una exactitud de 99.8 %, precisión de hasta 99.9 % y sensibilidad de 99.8 % para VGG16.

El trabajo de Xu et al. (2019) propuso una arquitectura personalizada compuesta por 28 capas, que logró una exactitud del 96.37 %, precisión del 78.11 % y sensibilidad del 100 %. En su análisis comparativo con modelos clásicos, VGG16 presentó un desempeño inferior con una exactitud de 84.67 %, precisión 45.77 % y sensibilidad 100 %; mientras que ResNet50 alcanzó una exactitud de 89.45 %, precisión del 55.09 % y sensibilidad del 100 %. Estos resultados son congruentes con los alcanzados en el presente trabajo de titulación en cuanto a sensibilidad, aunque muestran una caída considerable en precisión, posiblemente asociada al balance de clases o al tipo de preprocesamiento aplicado en dicha investigación. En contraste, este trabajo logró mantener un equilibrio adecuado entre precisión y sensibilidad, lo que refuerza la estabilidad de los modelos evaluados frente a bases de datos más realistas.

Alkannad et al. (2025) llevaron a cabo una evaluación comparativa de modelos preentrenados, entre ellos ResNet50, InceptionV3 y Xception, utilizando dos conjuntos de datos ampliamente reconocidos: METU y SDNET2018. Con el dataset METU, ResNet50

alcanzó resultados excepcionales, con una exactitud del 99.95 %, precisión del 99.9 % y una sensibilidad del 100 %, mientras que al aplicar el mismo modelo sobre el conjunto SDNET2018, los valores descendieron ligeramente, obteniendo una exactitud del 98.46 %, precisión del 99.45 % y sensibilidad del 97.45 % para imágenes de pavimentos.

De manera similar, Philip et al. (2023) reportaron desempeños sobresalientes: una exactitud de prueba de 99.61 % para VGG16, 99.67 % para VGG19 y 99.88 % para ResNet50, valores alcanzados en condiciones altamente controladas. Por su parte, Zoubir et al. (2022) emplearon VGG16 como base en tres enfoques distintos, logrando exactitudes de validación del 90.62 %, 94.80 % y 96.68 %, con precisiones de 83.04 %, 89.66 % y 94.89 %, y sensibilidades de 90.57 %, 98.86 % y 100 %, respectivamente. En una publicación anterior, Zoubir et al. (2021) obtuvieron resultados similares: una exactitud de validación de 97.50 % y de prueba de 94.89 % para VGG16, mientras que VGG19 logró 96.25 % en validación y 95.39 % en prueba, e InceptionV3 alcanzó 96.88 % y 95.89 %, respectivamente.

Cabe señalar que, algunas investigaciones reportaron únicamente la métrica de exactitud, mientras que este trabajo integró las tres métricas de exactitud, precisión y sensibilidad, lo cual proporciona una evaluación más integral del desempeño, especialmente útil en contextos con posibles desbalances entre clases. Además, aunque algunos estudios calcularon el F1-score como métrica adicional, en esta investigación se priorizó el análisis individual de precisión y sensibilidad, dado que se requería evaluar con mayor claridad la capacidad de los modelos para evitar falsos negativos, es decir, imágenes con fisuras clasificadas incorrectamente, que son una condición crítica en la inspección estructural.

Si bien los modelos entrenados en este trabajo de titulación no alcanzaron valores superiores al 99 % como en algunas investigaciones revisadas, los resultados obtenidos son altamente competitivos, considerando el tamaño de la base de datos de 8,000 imágenes, la

complejidad de las imágenes y la validación realizada sobre un entorno estructural real. En particular, la alta precisión de VGG16 (96.75 %) sugiere una baja tasa de falsos positivos, lo que es útil para reducir alarmas innecesarias, mientras que la elevada sensibilidad de ResNet50 (90.62 %) asegura una buena detección de fisuras presentes, minimizando el riesgo de omisiones.

Conclusiones

Frente al desafío que representa la detección precisa y oportuna de fisuras en estructuras de hormigón, este estudio desarrolló y evaluó modelos de redes neuronales convolucionales (CNN), específicamente VGG16 y ResNet50, para la detección automatizada de fisuras, utilizando imágenes reales capturadas en campo. La finalidad fue aportar una herramienta innovadora y adaptable a las condiciones reales del proceso de inspección en el sector de la ingeniería civil, que continúa enfrentando desafíos debido a la dependencia de métodos tradicionales, mayormente manuales, que pueden ser lentos, subjetivos y propensos a errores humanos, especialmente al evaluar estructuras de difícil acceso o con deterioro emergente.

El primer objetivo específico de esta investigación fue alcanzado con éxito mediante la elaboración de base de datos original compuesta por 8,000 imágenes de estructuras de hormigón en condiciones reales, capturadas con una cámara profesional y un dron, para la retroalimentación de los modelos CNN. Esta base de datos representa un aporte significativo, ya que se aleja de los entornos controlados de las bases de datos públicas comúnmente utilizadas. La calidad y variabilidad inherentes a las imágenes obtenidas en campo reflejan escenarios reales de inspección, lo que añade complejidad al problema pero también eleva el valor práctico de los modelos entrenados.

Además, la decisión de dividir el conjunto de datos exclusivamente en entrenamiento (75 %) y validación (25 %), prescindiendo de un conjunto de prueba, permitió evaluar el desempeño de los modelos sobre un subconjunto exigente: imágenes tomadas únicamente del Puente del Redondel San Juan, una estructura real no intervenida. Esta estrategia metodológica aportó rigor a la validación, permitiendo medir con mayor realismo la capacidad del modelo para generalizar su aprendizaje fuera del entorno de entrenamiento.

Respecto al segundo objetivo, que consistió en desarrollar y evaluar modelos de CNN para la detección automatizada de fisuras, se logró implementar y comparar dos arquitecturas: VGG16 y ResNet50. Los resultados obtenidos reflejan un desempeño satisfactorio en ambos modelos. VGG16 alcanzó una exactitud del 88.95 %, una precisión del 96.75 % y una sensibilidad del 84.42 %. Estos valores indican que, si bien este modelo fue muy eficaz en evitar falsos positivos, tuvo más dificultades para identificar todos los casos positivos, es decir, aquellas imágenes que efectivamente contenían fisuras. Por su parte, ResNet50 mostró un rendimiento más equilibrado, con una exactitud del 90.80 %, precisión del 86.93 % y sensibilidad del 90.62 %. Su mayor sensibilidad sugiere una mejor capacidad para detectar fisuras, aunque con una ligera tendencia a clasificar erróneamente algunas imágenes sin fisuras como defectuosas. En ambos casos, los resultados reflejan una capacidad destacada para resolver la tarea planteada, en un contexto donde la detección automatizada se dificulta por la naturaleza no estandarizada de los datos.

En comparación con otros estudios similares, los resultados obtenidos en este trabajo de titulación se sitúan dentro de un rango competitivo, aunque ligeramente por debajo de aquellos alcanzados en investigaciones que utilizaron bases de datos públicas limpias y balanceadas como SDNET2018 o METU. Sin embargo, es importante destacar que estas investigaciones operan en condiciones ideales o semi-controladas, lo que facilita la obtención de métricas más elevadas. La presente investigación aporta valor precisamente por demostrar que es posible alcanzar altos niveles de precisión y sensibilidad utilizando datos reales y no tratados.

Esta investigación contribuye al fortalecimiento del diagnóstico estructural en la ingeniería civil, proporcionando una solución innovadora basada en inteligencia artificial que responde a la creciente necesidad de automatizar procesos de inspección y monitoreo estructural. Al desarrollar y validar modelos de redes neuronales convolucionales aplicados a

imágenes reales de estructuras de hormigón, se propone una alternativa innovadora a los métodos tradicionales de inspección visual. Esta iniciativa fortalece la adopción de tecnologías emergentes en una disciplina que enfrenta desafíos crecientes en la gestión de infraestructuras envejecidas, el mantenimiento preventivo y la garantía de seguridad estructural. Al demostrar la viabilidad técnica del uso de modelos como VGG16 y ResNet50 en condiciones reales, el estudio sienta las bases para una integración más amplia de soluciones automatizadas dentro de los procesos de diagnóstico estructural.

Cabe destacar que, aunque los modelos CNN aplicados en este estudio demostraron resultados prometedores en la detección de fisuras, aún existen oportunidades de mejora para futuras investigaciones. Se recomienda ampliar la base de datos con imágenes más diversas y representativas de una mayor variedad de estructuras, niveles de deterioro y condiciones climáticas. Además, se sugiere explorar arquitecturas más modernas como EfficientNet y MobileNetV3, que podrían ofrecer una mejora tanto en la precisión como en la eficiencia del procesamiento. También resulta relevante considerar el uso de equipos con mejor capacidad técnica y computacional para optimizar el entrenamiento de modelos más complejos. Como línea futura de investigación, se propone la incorporación de tareas complementarias como la segmentación de fisuras y estimación de su ancho, implementando sistemas automatizados en tiempo real, como drones equipados con cámaras de alta resolución y algoritmos embebidos que permitan identificar y medir daños automáticamente.

En general, este trabajo de investigación demuestra que la combinación entre visión por computadora, redes neuronales profundas y datos reales capturados en campo tiene el potencial de transformar la forma en la que se inspeccionan estructuras de hormigón, abriendo nuevas posibilidades para desarrollar soluciones más rápidas, objetivas y automatizadas, contribuyendo así a una gestión más eficiente y al avance tecnológico de la ingeniería civil.

Referencias Bibliográficas

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Brain, G. (2016). *TensorFlow: A system for large-scale machine learning*. <https://export.arxiv.org/pdf/1605.08695v2.pdf>
- Ai, D., Jiang, G., Lam, S. K., He, P., & Li, C. (2023). Computer vision framework for crack detection of civil infrastructure—A review. *Engineering Applications of Artificial Intelligence*, 117, 105478. <https://doi.org/10.1016/J.ENGAPPAL.2022.105478>
- Ali, L., Alnajjar, F., Jassmi, H. Al, Gochoo, M., Khan, W., & Serhani, M. A. (2021). Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures. *Sensors 2021*, Vol. 21, Page 1688, 21(5), 1688. <https://doi.org/10.3390/S21051688>
- Alkannad, A. A., Smadi, A. Al, Yang, S., Al-Smadi, M. K., Al-Makhlafi, M., Feng, Z., & Yin, Z. (2025). CrackVision: Effective Concrete Crack Detection with Deep learning and Transfer Learning. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3540841>
- American Concrete Institute. (2001). *Control of Cracking in Concrete Structures (ACI 224R-01)*.
- American Concrete Institute. (2023). *What an owner should know about nondestructive testing—TechNote. (ACI PRC-228.3-23)*.
- Aslan, O. D., Gultepe, E., Ramaji, I. J., & Kermanshachi, S. (2019). Using artificial intelligence for automating pavement condition assessment. *International Conference on Smart Infrastructure and Construction 2019, ICSIC 2019: Driving Data-Informed Decision-Making*, 337–341.

<https://doi.org/10.1680/ICSIC.64669.337/ASSET/IMAGES/SMALL/ICSIC.64669.337>.

F4.GIF

Billah, U. H., Tavakkoli, A., & La, H. M. (2019). Concrete Crack Pixel Classification Using an Encoder Decoder Based Deep Learning Architecture. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11844 LNCS, 593–604. https://doi.org/10.1007/978-3-030-33720-9_46

Calavera, J. (2005). *Patología de Estructuras de Hormigón Armado y Pretensado Tomo I* (2da ed.). INTEMAC.

Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, 32(5), 361–378.

<https://doi.org/10.1111/MICE.12263>;JOURNAL:JOURNAL:14678667;WGROU:STR
ING:PUBLICATION

Chen, B., Zhang, H., Chen, Y., Li, Y., & Xiong, J. (2023). Semantic segmentation method of hydraulic structure crack based on feature enhancement. *Qinghua Daxue Xuebao/Journal of Tsinghua University*, 63(7), 1135–1143.

<https://doi.org/10.16511/J.CNKI.QHDXXB.2023.26.009>

Chen, K., Yadav, A., Khan, A., Meng, Y., & Zhu, K. (2019). Improved Crack Detection and Recognition Based on Convolutional Neural Network. *Modelling and Simulation in Engineering*, 2019. <https://doi.org/10.1155/2019/8796743>

Chen, L., Chen, W., Wang, L., Zhai, C., Hu, X., Sun, L., Tian, Y., Huang, X., & Jiang, L. (2023). Convolutional neural networks (CNNs)-based multi-category damage detection and recognition of high-speed rail (HSR) reinforced concrete (RC) bridges using test

images. *Engineering Structures*, 276, 115306.

<https://doi.org/10.1016/J.ENGSTRUCT.2022.115306>

Choi, Y., Bae, B., Han, T. H., & Ahn, J. (2024). Application of Mask R-CNN and YOLOv8 Algorithms for Concrete Crack Detection. *IEEE Access*.

<https://doi.org/10.1109/ACCESS.2024.3469951>

Choi, Y., Park, H. W., Mi, Y., & Song, S. (2024). Crack Detection and Analysis of Concrete Structures Based on Neural Network and Clustering. *Sensors* 2024, Vol. 24, Page 1725, 24(6), 1725. <https://doi.org/10.3390/S24061725>

Deng, J., Singh, A., Zhou, Y., Lu, Y., & Lee, V. C. S. (2022). Review on computer vision-based crack detection and quantification methodologies for civil structures. *Construction and Building Materials*, 356, 129238.

<https://doi.org/10.1016/J.CONBUILDMAT.2022.129238>

Deng, Z., Yan, X., Zhang, S., & Bailey, C. P. (2020). *Extremal Region Analysis based Deep Learning Framework for Detecting Defects*. <https://arxiv.org/abs/2003.08525v2>

DJI. (s.f.). *DJI Phantom 4 RTK [Fotografia]*. DroneDepot:

<https://www.dronedepot.be/product/dji-phantom-4-rtk/>

Dorafshan, S., Thomas, R. J., & Maguire, M. (2018). SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks. *Data in Brief*, 21, 1664–1668. <https://doi.org/10.1016/J.DIB.2018.11.015>

Dyer, T. (2014). *Concrete Durability*. CRC Press. Taylor & Francis Group.

Fu, R., Xu, H., Wang, Z., Shen, L., Cao, M., Liu, T., & Novák, D. (2020). Enhanced Intelligent Identification of Concrete Cracks Using Multi-Layered Image Preprocessing-

- Aided Convolutional Neural Networks. *Sensors* 2020, Vol. 20, Page 2021, 20(7), 2021.
<https://doi.org/10.3390/S20072021>
- Gjørsv, O. E. (2011). Durability of concrete structures. *Arabian Journal for Science and Engineering*, 36(2), 151–172. <https://doi.org/10.1007/S13369-010-0033-5/METRICS>
- Goldsborough, P. (2016). *A Tour of TensorFlow*.
<https://export.arxiv.org/pdf/1610.01178v1.pdf>
- Guerri, M. F., Distante, C., Spagnolo, P., Bougourzi, F., & Taleb-Ahmed, A. (2024). Deep learning techniques for hyperspectral image analysis in agriculture: A review. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 12, 100062.
<https://doi.org/10.1016/J.OPHOTO.2024.100062>
- Guo, L., Li, R., Jiang, B., & Shen, X. (2020). Automatic crack distress classification from concrete surface images using a novel deep-width network architecture.
Neurocomputing, 397, 383–392. <https://doi.org/10.1016/J.NEUCOM.2019.08.107>
- Hui, L., Ibrahim, A., & Hindi, R. (2025). Computer Vision-Based Concrete Crack Identification Using MobileNetV2 Neural Network and Adaptive Thresholding.
Infrastructures 2025, Vol. 10, Page 42, 10(2), 42.
<https://doi.org/10.3390/INFRASTRUCTURES10020042>
- Hunt, J. (2019). *Introduction to Matplotlib*. 35–42. https://doi.org/10.1007/978-3-030-25943-3_5
- Infrastructure and Projects Authority. (2017). *Transforming Infrastructure Performance*.
www.gov.uk/IPA
- JetBrains. (2020, August). *Logo of PyCharm [Fotografia]*. JetBrains: .
<https://www.jetbrains.com/company/brand/#logos-and-icons-jetbrains-logos>

- Junior, W. M. P., da Silva, S. F., Silva, A. R. E., Rezio, L. H. F., da Silva, M. P., Guimarães, N. R. da S., & Canuto, S. D. C. (2024). Avaliação da presença de fissuras em imagens de estruturas de concreto através do uso de redes neurais profundas. *Matéria (Rio de Janeiro)*, 29(4), e20240354. <https://doi.org/10.1590/1517-7076-RMAT-2024-0354>
- Khandelwal, V. (2020). *The Architecture and Implementation of VGG-16 [Fotografia]*. Medium: <https://pub.towardsai.net/the-architecture-and-implementation-of-vgg-16-b050e5a5920b>
- Kim, B., Yuvaraj, N., Sri Preethaa, K. R., & Arun Pandian, R. (2021). Surface crack detection using deep learning with shallow CNN architecture for enhanced computation. *Neural Computing and Applications*, 33(15), 9289–9305. <https://doi.org/10.1007/s00521-021-05690-8>
- Koch, C., Paal, S., Rashidi, A., Zhu, Z., König, M., & Brilakis, I. (2014). Achievements and Challenges in Machine Vision-Based Inspection of Large Concrete Structures. *Advances in Structural Engineering*, 17(3), 303–318. <https://doi.org/10.1260/1369-4332.17.3.303>
- Konovalenko, I., Maruschak, P., Brezinová, J., Viňáš, J., & Brezina, J. (2020). Steel surface defect classification using deep residual neural network. *Metals*, 10(6), 1–15. <https://doi.org/10.3390/met10060846>
- Lee, S. H., Kim, H. Y., Shin, H. K., Jang, Y., & Ahn, Y. H. (2017). Introducing a model for evaluating concrete structure performance using deep convolutional neural network. *International Journal of Sustainable Building Technology and Urban Development* 2017 8:3, 8(3), 285–295. <https://doi.org/10.12972/SUSB.20170027>
- Lenovo. (s.f.). *Lenovo IdeaPad 3 Notebook 15,6" [Fotografia]*. Amazon: <https://www.amazon.es/Lenovo-IdeaPad-Notebook-procesador-reconocimiento/dp/B09BNQ618C>

- Li, L. F., Ma, W. F., Li, L., & Lu, C. (2019). Research on Detection Algorithm for Bridge Cracks Based on Deep Learning. *Zidonghua Xuebao/Acta Automatica Sinica*, 45(9), 1727–1742. <https://doi.org/10.16383/J.AAS.2018.C170052>
- Luo, L. (2021). Research on Image Classification Algorithm Based on Convolutional Neural Network. *Journal of Physics: Conference Series*, 2083(3). <https://doi.org/10.1088/1742-6596/2083/3/032054>
- Ma, D., Fang, H., Wang, N., Xue, B., Dong, J., & Wang, F. (2021). A real-time crack detection algorithm for pavement based on CNN with multiple feature layers. *Road Materials and Pavement Design*, 23(9), 2115–2131. <https://doi.org/10.1080/14680629.2021.1925578>;CTYPE:STRING:JOURNAL
- Mayya, A., Alkayem, N. F., Shen, L., Zhang, X., Fu, R., Wang, Q., & Cao, M. (2024). Efficient hybrid ensembles of CNNs and transfer learning models for bridge deck image-based crack detection. *Structures*, 64, 106538. <https://doi.org/10.1016/J.ISTRUC.2024.106538>
- McKinsey & Company. (2017). *Reinventing Construction: A Route to Higher Productivity*.
- Modarres, C., Astorga, N., Droguett, E. L., & Meruane, V. (2018). Convolutional neural networks for automated damage recognition and damage type identification. *Structural Control and Health Monitoring*, 25(10), e2230. <https://doi.org/10.1002/STC.2230>;PAGEGROUP:STRING:PUBLICATION
- Nelli, F. (2015). The NumPy Library. *Python Data Analytics*, 35–61. https://doi.org/10.1007/978-1-4842-0958-5_3
- Nugraheni, D. M. K., Nugroho, A. K., Dewi, D. I. K., & Noranita, B. (2023). Deca Convolutional Layer Neural Network (DCL-NN) Method for Categorizing Concrete

- Cracks in Heritage Building. *International Journal of Advanced Computer Science and Applications*, 14(1), 722–730. <https://doi.org/10.14569/IJACSA.2023.0140180>
- Ojeda, J. M. P., Cayatopa-Calderón, B. A., Huatangari, L. Q., Tineo, J. L. P., Pino, M. E. M., & Pintado, W. R. (2023). Convolutional Neural Network for Predicting Failure Type in Concrete Cylinders During Compression Testing. *Civil Engineering Journal*, 9(9), 2105–2119. <https://doi.org/10.28991/CEJ-2023-09-09-01>
- Özgenel, Ç. F., & Gönenç Sorguç, A. (2019). *Concrete Crack Images for Classification*. 2. <https://doi.org/10.17632/5Y9WDSG2ZT.2>
- Pang, B., Nijkamp, E., & Wu, Y. N. (2020). Deep Learning With TensorFlow: A Review. *Journal of Educational and Behavioral Statistics*, 45(2), 227–248. <https://doi.org/10.3102/1076998619872761;CTYPE:STRING:JOURNAL>
- Philip, R. E., Andrushia, A. D., Nammalvar, A., Gurupatham, B. G. A., & Roy, K. (2023). A Comparative Study on Crack Detection in Concrete Walls Using Transfer Learning Techniques. *Journal of Composites Science* 2023, Vol. 7, Page 169, 7(4), 169. <https://doi.org/10.3390/JCS7040169>
- Plotka, S., Włodarczyk, T., Szczerba, R., Chomiak, A., & Komisarek, O. (2021). Machine learning methods as an aid in planning orthodontic treatment on the example of Cone-Beam Computed Tomography analysis: a literature review. *Journal of Education, Health and Sport*, 11(1), 94–104. <https://doi.org/10.12775/JEHS.2021.11.01.010>
- Prakash, K. B., Ruwali, A., & Kanagachidambaresan, G. R. (2021). Introduction to Tensorflow Package. *EAI/Springer Innovations in Communication and Computing*, 1–4. https://doi.org/10.1007/978-3-030-57077-4_1

- Quqa, S., Martakis, P., Movsessian, A., Pai, S., Reuland, Y., & Chatzi, E. (2022). Two-step approach for fatigue crack detection in steel bridges using convolutional neural networks. *Journal of Civil Structural Health Monitoring*, 12(1), 127–140.
<https://doi.org/10.1007/S13349-021-00537-1/METRICS>
- Rajagopalan, G. (2021). Working with NumPy Arrays. *A Python Data Analyst's Toolkit*, 117–145. https://doi.org/10.1007/978-1-4842-6399-0_5
- Ramazhan, M. R. S., Bustamam, A., & Buyung, R. A. (2025). Smart Car Damage Assessment Using Enhanced YOLO Algorithm and Image Processing Techniques. *Information 2025*, Vol. 16, Page 211, 16(3), 211.
<https://doi.org/10.3390/INFO16030211>
- Rashid, T., Mokji, M. M., & Rasheed, M. (2024). Cracked concrete surface classification in low-resolution images using a convolutional neural network. *Journal of Optics (India)*, 1–13. <https://doi.org/10.1007/S12596-024-02080-W/METRICS>
- Sapre, A., & Vartak, S. (2020). Scientific Computing and Data Analysis using NumPy and Pandas. *Inter International Research Journal of Engineering and Technology*.
www.irjet.net
- Sargiotis, D. (2024). *Advancing Civil Engineering with AI and Machine Learning: From Structural Health to Sustainable Development*. <https://doi.org/10.2139/SSRN.4883999>
- Shafiei Dizaji, M., & Harris, D. (2019). 3D InspectionNet: a deep 3D convolutional neural networks based approach for 3D defect detection on concrete columns.
<https://doi.org/10.1117/12.2514387>, 10971, 67–77.
<https://doi.org/10.1117/12.2514387>

SONY ELECTRONICS INC. (s.f.-a). *Alpha 6100 APS-C Mirrorless Interchangeable-Lens Camera + 16-50mm Lens [Fotografia]*. SONY:

<https://electronics.sony.com/imaging/interchangeable-lens-cameras/aps-c/p/ilce6100kb>

SONY ELECTRONICS INC. (s.f.-b). *Sony Alpha 6100 APS-C Mirrorless Interchangeable-Lens Camera + 16-50mm Lens | ILCE-6100K*.

<https://electronics.sony.com/imaging/interchangeable-lens-cameras/aps-c/p/ilce6100kb>

Tello-Gil, C., Jabari, S., Waugh, L., Masry, M., & McGinn, J. (2025). Improving Crack Detection on Concrete Structures Using Real-World Data Augmentation for Deep Learning Convolutional Neural Networks. *Lecture Notes in Civil Engineering*, 498 LNCE, 237–249. https://doi.org/10.1007/978-3-031-61499-6_18

Wen, W., Xu, T., Hu, J., Ji, D., Yue, Y., & Zhai, C. (2025). Seismic damage recognition of structural and non-structural components based on convolutional neural networks. *Journal of Building Engineering*, 102, 112012. <https://doi.org/10.1016/J.JOBE.2025.112012>

World Economic Forum. (2016). *Shaping the Future of Construction A Breakthrough in Mindset and Technology Prepared in collaboration with The Boston Consulting Group*.

Wu, Y., Li, S., Li, J., Yu, Y., Li, J., & Li, Y. (2025). Deep learning in crack detection: A comprehensive scientometric review. *Journal of Infrastructure Intelligence and Resilience*, 4(3), 100144. <https://doi.org/10.1016/J.IINTEL.2025.100144>

Xu, H., Su, X., Wang, Y., Cai, H., Cui, K., & Chen, X. (2019). Automatic Bridge Crack Detection Using a Convolutional Neural Network. *Applied Sciences* 2019, Vol. 9, Page 2867, 9(14), 2867. <https://doi.org/10.3390/APP9142867>

- Yabi, C. P., Gbehoun, G. F., Tamou, B. C. K., Alamou, E., Gibigaye, M., & Farsangi, E. N. (2025). Analysis and Classification of Distress on Flexible Pavements Using Convolutional Neural Networks: A Case Study in Benin Republic. *Infrastructures* 2025, Vol. 10, Page 111, 10(5), 111. <https://doi.org/10.3390/INFRASTRUCTURES10050111>
- Yamane, T., & Chun, P. J. (2020). Crack Detection from a Concrete Surface Image Based on Semantic Segmentation Using Deep Learning. *Journal of Advanced Concrete Technology*, 18(9), 493–504. <https://doi.org/10.3151/JACT.18.493>
- Yang, L., Ma, Z., Zhu, L., -, al, Balaji, N., Karthik Pai, B. H., Bhat, B., Cao, S., Yang, S., & Cao, S. (2021). Research on Python Data Visualization Technology. *Journal of Physics: Conference Series*, 1757(1), 012122. <https://doi.org/10.1088/1742-6596/1757/1/012122>
- Yao, G., Sun, Y., Wong, M., Lv, X., Yang, Y., Lei, Y., Meng, X., & Li, J. (2021). A Real-Time Detection Method for Concrete Surface Cracks Based on Improved YOLOv4. *Symmetry* 2021, Vol. 13, Page 1716, 13(9), 1716. <https://doi.org/10.3390/SYM13091716>
- Zadeh, S. S., Birgani, S. A., Khorshidi, M., & Kooban, F. (2023). Concrete Surface Crack Detection with Convolutional-based Deep Learning Models. *International Journal of Novel Research in Civil Structural and Earth Sciences*, 10, 25–35. <https://doi.org/10.5281/zenodo.10061654>
- Zhang, L., Yang, F., Daniel Zhang, Y., & Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. *Proceedings - International Conference on Image Processing, ICIP, 2016-August*, 3708–3712. <https://doi.org/10.1109/ICIP.2016.7533052>
- Zhu, J., & Song, J. (2020). An Intelligent Classification Model for Surface Defects on Cement Concrete Bridges. *Applied Sciences* 2020, Vol. 10, Page 972, 10(3), 972. <https://doi.org/10.3390/APP10030972>

- Zoubir, H., Rguig, M., El Aroussi, M., Chehri, A., Saadane, R., & Jeon, G. (2022). Concrete Bridge Defects Identification and Localization Based on Classification Deep Convolutional Neural Networks and Transfer Learning. *Remote Sensing 2022, Vol. 14, Page 4882, 14*(19), 4882. <https://doi.org/10.3390/RS14194882>
- Zoubir, H., Rguig, M., El Aroussi, M., Saadane, R., & Chehri, A. (2024). Pixel-level concrete bridge crack detection using Convolutional Neural Networks, gabor filters, and attention mechanisms. *Engineering Structures, 314*, 118343. <https://doi.org/10.1016/J.ENGSTRUCT.2024.118343>
- Zoubir, H., Rguig, M., & Elaroussi, M. (2021). Crack recognition automation in concrete bridges using Deep Convolutional Neural Networks. *MATEC Web of Conferences, 349*, 03014. <https://doi.org/10.1051/matecconf/202134903014>