

Evaluación de precisión en modelos de visión por computadora basados en Deep Learning para la detección de grietas en pavimentos flexibles en la localidad de Puente Aranda, Bogotá

Jonathan Steven Olivos Fonseca

José David Muñoz Gutiérrez

Kevin Steven Morgado Gómez

Asesor

Yimy Alejandro Bernal Campos

Universidad EAN

Facultad de Ingeniería

Seminario de investigación

2026

Tabla de Contenido

Introducción	7
Planteamiento del problema.....	8
Pregunta de investigación	8
Objetivos.....	9
Justificación	10
Antecedentes Investigativos.....	13
Marco normativo e institucional	22
Marco teórico	25
Metodología	32
Enfoque metodológico	32
Variables técnicas y contextuales	33
Método de recolección de información	34
Proceso de etiquetado para entrenamiento	38
Localización de las imágenes capturadas.....	40
Parámetros Modelo Deep Learning	46
Resultados.....	51
Dataset de referencia	51
Etiquetado	52
Entrenamiento modelo base	55
Gráficas de desempeño.....	56

Validación con imágenes de grietas.....	60
Comparación con investigaciones anteriores.....	62
Conclusiones	67
Referencias	69

Figuras

Figura 1. Tipos de Grietas en Pavimento	14
Figura 2. Esquema de desarrollo de detección de grietas	15
Figura 3. <i>Estructura de U-Net</i>	15
Figura 4. Desarrollo de análisis de grietas con modelos XGBoost.....	17
Figura 5. Ejemplo de identificación de grietas por diferentes modelos	18
Figura 6. Ejemplo de análisis de imágenes de grietas.....	19
Figura 7. Topología de red Yolo-LY.....	20
Figura 8. Esquema modelo Yolo V5	21
Figura 9. Esquema de Árbol de decisión	25
Figura 10. Muestra del proceso de entrenamiento del XGBoost.....	26
Figura 11. Arquitectura Red Neuronal	27
Figura 12. Esquema de intersección sobre Unión	29
Figura 13. Desempeño de modelos de referencia YOLO	30
Figura 14. Localización general — Tramos evaluados sobre Av. 1 de Mayo y Transversal 53, Bogotá D.C.	41
Figura 15. Localización Tramo 1 — Avenida 1 de Mayo.	44
Figura 16. Localización Tramo 2 — Calle 19 Sur.....	44
Figura 17. Localización Tramo 3 — Carrera 52	45
Figura 18. Localización Tramo 4 — Transversal 53.....	46
Figura 19. Diagrama de flujo del proyecto	50
Figura 20. Muestra de grietas en set de entrenamiento.....	51
Figura 21. Interfaz de etiquetado de imágenes.....	52
Figura 22. Muestra de grietas #1.....	53
Figura 23. Muestra de grietas #2.....	53
Figura 24. Muestra de grietas #3.....	54

Figura 25. Etiquetado de grietas en pavimento en Label-studio	54
Figura 26. Muestra de grietas de entrenamiento	55
Figura 27. Gráfica de pérdida.....	57
Figura 28. Evolución de precisión.....	57
Figura 29. Gráfica de evolución de recall	58
Figura 30. Métrica mAP50-95.....	58
Figura 31. Ejemplo de clasificación de grietas.....	59
Figura 32. Desempeño en conjunto de imágenes de entrenamiento	59
Figura 33. Matriz de confusión entrenamiento.....	60
Figura 34. Muestra de marcación real	61
Figura 35. Predicciones del modelo Yolo en Grietas.....	61
Figura 36. Predicciones del modelo Yolo en Grietas	62
Figura 37. Muestras de grietas en entrenamiento - comparativa	62
Figura 38. Muestra de predicción de grietas verticales.....	63
Figura 39. Muestra de predicción de grietas horizontales y verticales	63
Figura 40. Matriz de confusión predicción de grietas.....	65

Tablas

Tabla 1. Descripción de variables técnicas que afectan el análisis de grietas en pavimento flexible	33
Tabla 2. Descripción de variables contextuales que afectan el análisis de grietas en pavimento flexible	34
Tabla 3. Resumen de características imágenes de predicción de grietas	35
Tabla 4. Convención cromática de deterioro por rango PCI — ASTM D6433.....	43
Tabla 5. Resumen de parámetros Red YOLO V26 nano.....	47
Tabla 6. Valores esperados de desempeño	49
Tabla 7. Comparación de métricas para entrenamiento	64
Tabla 8. Comparación de métricas para validación.....	64

Introducción

El deterioro de los pavimentos flexibles, especialmente en forma de grietas, es un problema frecuente en la infraestructura vial y afecta tanto la seguridad como la planificación del mantenimiento. Tradicionalmente, la identificación de estas fallas se realiza mediante inspección visual, un proceso que depende en gran medida del criterio del evaluador y que puede presentar variaciones en los resultados. Frente a esto, en los últimos años han surgido alternativas basadas en visión por computadora y Deep Learning, que permiten automatizar la detección de grietas y analizar imágenes de forma más consistente.

En este trabajo se realizará una evaluación de la precisión de un modelo de detección de grietas aplicado a imágenes de pavimentos flexibles, se compararán sus resultados con los obtenidos mediante inspección visual tradicional. Para ello, se utilizarán imágenes capturadas en campo, sobre las cuales se aplicará el modelo y se identificarán las grietas detectadas. Estos resultados se contrastarán con una referencia construida a partir de la inspección visual, con el fin de analizar cuan cercano será el desempeño del modelo frente al método convencional.

El alcance del trabajo se centrará en comparar ambos enfoques desde un punto de vista cuantitativo, utilizando métricas como precisión, recall y F1-score. No se busca desarrollar un nuevo modelo, sino evaluar el comportamiento de una herramienta existente en un contexto específico. Con esto, se pretende aportar una visión más clara sobre el uso de este tipo de tecnologías como apoyo en la detección de grietas en pavimentos flexibles.

Planteamiento del problema

El levantamiento de daños de tipo grietas superficiales en pavimentos flexibles constituye una actividad fundamental para el diagnóstico del estado de la infraestructura vial y la toma de decisiones en los procesos de mantenimiento y rehabilitación. Tradicionalmente, este levantamiento se realiza mediante inspección visual en campo, apoyada en formatos normalizados, lo cual implica altos tiempos de ejecución, costos operativos elevados y un componente significativo de subjetividad asociado a la experiencia del evaluador.

En el contexto actual, las entidades encargadas de la gestión vial como el Instituto de Desarrollo Urbano de Bogotá (IDU) enfrentan restricciones presupuestales y la necesidad de optimizar recursos, lo que exige métodos más eficientes y confiables para la recolección de información, por ejemplo, en las vías de la localidad de Puente Aranda. En este escenario, el uso de herramientas basadas en inteligencia artificial para el reconocimiento automático de fallas superficiales surge como una alternativa tecnológica con potencial para mejorar la exactitud del diagnóstico, reducir tiempos de levantamiento y disminuir costos operativos.

No obstante, la adopción de estas tecnologías requiere ser analizada comparando su nivel de precisión frente a la inspección visual tradicional, con el fin de establecer si realmente representan una mejora en la identificación de daños superficiales. En este sentido, surge la necesidad de examinar la evidencia disponible respecto al desempeño de los modelos basados en Deep Learning en comparación con los métodos convencionales.

Pregunta de investigación

¿En qué medida los modelos de visión por computadora basados en Deep Learning mejoran la precisión en la detección de grietas en pavimentos flexibles en comparación con la inspección visual tradicional?

Objetivos

Objetivo general

Evaluar la precisión de los modelos de visión por computadora basados en Deep Learning, en comparación con la inspección visual tradicional para la identificación de grietas en pavimentos flexibles en la localidad de Puente Aranda, Bogotá.

Objetivos específicos

- Identificar los principales modelos de visión por computadora basados en Deep Learning utilizados en la detección de grietas en pavimentos flexibles.
- Determinar las métricas empleadas en estudios relevantes para evaluar la precisión en la detección de grietas tanto en modelos basados en Deep Learning como en la inspección visual tradicional.
- Definir y realizar el etiquetado manual de grietas en imágenes de pavimentos flexibles, con el fin de apoyar el entrenamiento del modelo de detección y construir un conjunto de referencia que permita comparar sus resultados con la inspección visual tradicional.
- Comparar la precisión obtenida mediante la ejecución simulada de un modelo basado en Deep Learning con la de la inspección visual tradicional en la detección de grietas en pavimentos flexibles.

Justificación

El levantamiento de daños superficiales en pavimentos flexibles constituye un componente esencial dentro de los sistemas de gestión de infraestructura vial, ya que la calidad del diagnóstico condiciona la priorización de intervenciones, la asignación presupuestal y la planificación del mantenimiento. Tradicionalmente, este proceso se realiza mediante inspección visual en campo; sin embargo, dicha metodología presenta limitaciones asociadas a la subjetividad del evaluador y la dependencia de la experiencia individual, factores que pueden afectar la consistencia del inventario de deterioros.

En los últimos años, la evolución de técnicas de visión por computador y *Deep Learning* ha transformado el campo de la detección automática de fallas. La literatura reciente evidencia una transición tecnológica desde métodos clásicos de procesamiento de imágenes hacia arquitecturas basadas en redes neuronales profundas, clasificadas en enfoques de clasificación, detección de objetos y segmentación semántica (Hemdan & Al-Atroush, 2025). Dentro de estas, la segmentación a nivel pixel ha demostrado mayor capacidad para representar la geometría real de las grietas, aunque enfrenta desafíos relacionados con el desbalance entre píxeles de grieta y fondo, variabilidad en iluminación y presencia de anomalías superficiales.

Estudios recientes reportan avances significativos en precisión experimental. El modelo ShuttleCrackNet, basado en una arquitectura híbrida que integra mecanismos de atención y módulos tipo Transformer, alcanza un F-measure (precisión y sensibilidad) de 91,97 % y un IoU (qué tanto coincide el área detectada por el modelo con el área real de la grieta.) de 85,56 % en pruebas sobre 1500 imágenes (Zhang et al., 2025). Estos resultados superan arquitecturas ampliamente utilizadas como U-Net, SegNet y DeepLabV3+, particularmente en la continuidad de grietas finas y en la reducción de segmentaciones fragmentadas.

No obstante, la detección a nivel pixel es altamente sensible a variaciones ambientales como cambios en las condiciones de iluminación, presencia de sombras proyectadas, heterogeneidad del material asfáltico, humedad superficial y la existencia de elementos ajenos a la fisuración —tales como marcas viales, sellos o parches—. Estos factores modifican el contraste, la reflectancia y la textura de la imagen, dificultando la discriminación precisa entre píxeles correspondientes a grietas y píxeles de fondo.

En consecuencia, la diversidad de materiales y la presencia de objetos no relacionados con el deterioro pueden inducir errores en el registro de las imágenes, lo que implica que pueden generarse errores de clasificación, ya sea identificando como grietas elementos que no lo son (falsos positivos) o dejando de detectar fisuras reales presentes en la superficie (falsos negativos).

Desde una perspectiva más amplia, revisiones sistemáticas recientes en el ámbito de la detección automática de patologías superficiales en un pavimento flexible evidencian que, aunque el desempeño de los modelos ha mejorado sustancialmente, persisten limitaciones relacionadas con la generalización a diferentes escenarios operativos, la dependencia de bases de datos etiquetadas y la transferencia efectiva a contextos reales de gestión vial (Zhang et al., 2026). Esta brecha entre desempeño experimental y aplicabilidad práctica constituye un vacío investigativo relevante.

En consecuencia, la presente investigación se justifica, en primer lugar, desde una perspectiva técnica, al analizar la evidencia científica disponible sobre la precisión reportada por los modelos de visión por computadora basados en Deep Learning y la inspección visual tradicional en la detección de grietas en pavimentos flexibles, a partir de métricas estandarizadas utilizadas en estudios previos.

En segundo lugar, desde una perspectiva científica, el estudio busca organizar y comparar los resultados existentes en la literatura, con el fin de identificar las principales fortalezas y limitaciones de los enfoques automatizados frente a los métodos convencionales.

De esta manera, la investigación contribuye a la discusión académica sobre el uso de tecnologías automatizadas en el diagnóstico de pavimentos, con base en resultados previamente reportados y verificables.

Antecedentes Investigativos

A continuación, se presentan investigaciones enfocadas en la detección de grietas en pavimentos, que representan la importancia de esta problemática junto con los avances a nivel científico de los modelos de Machine Learning y los data set para mejorar el diagnóstico, por medio de sistemas viables para mejorar los niveles de mantenimiento y seguridad en las vías.

Como punto de inicio se identifica la importancia que ha tomado la visión de computadora, los modelos de lenguaje natural y machine learning en el diagnóstico de estado de las vías. (Tafida et al., 2025; Zhang, Li, & Shen, 2026) Esto se genera por la necesidad de cambiar los métodos tradicionales de diagnóstico de calidad de las vías, iniciando por los altos costos de realización y del peso que tiene el conocimiento experto en dicha actividad, lo que conlleva a sesgos o resultados heterogéneos.

Por ello, se ha buscado incentivar la toma de decisiones basadas en datos con la generación de nuevas estrategias de mantenimiento, dentro de campos de la innovación como las ciudades y transporte inteligentes. Pero se identifica un reto con las imágenes que se utilizan para este diagnóstico y cuellos de botella que se generan por la limitación de los modelos de Machine Learning. (Zeng et al., 2022)

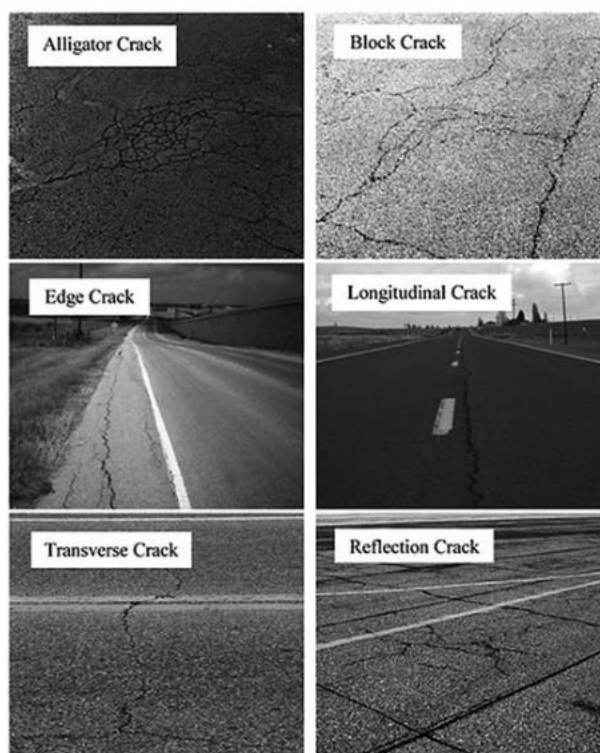
En (El-Din Hemdan & Al-Atroush, 2025) se hace la revisión de las tecnologías utilizadas en la detección de grietas y se identificó la dirección de la innovación por medio de metodologías de Deep Learning y el Internet de las Cosas. Puntualmente, se consideró la limitación de los enfoques tradicionales basados en la precisión de detección de fallas, automatización del flujo de trabajo y el manejo de grandes volúmenes de información utilizando Inteligencia Artificial. Dentro del campo de estudio se consideran los siguientes tipos de grietas, complementados con la Figura 1:

- **Grietas de cocodrilo o de fatiga:** causadas por el tráfico y que indican una falla estructural del pavimento.

- **Grietas de bloque:** Forman un patrón de bloque en la superficie del pavimento, asociada con cambios en la temperatura y contracción del asfalto a lo largo del tiempo.
- **Grietas en los bordes:** Ocurren cerca de los bordes, causadas por condiciones del ambiente como la temperatura y mal drenaje.
- **Grietas longitudinales:** Perpendiculares al centro del pavimento por cambios de temperatura.
- **Grietas de reflexión:** Causadas por capas de asfalto construidas sobre pavimentos de concreto por movimiento o expansión térmica.

Figura 1.

Tipos de Grietas en Pavimento



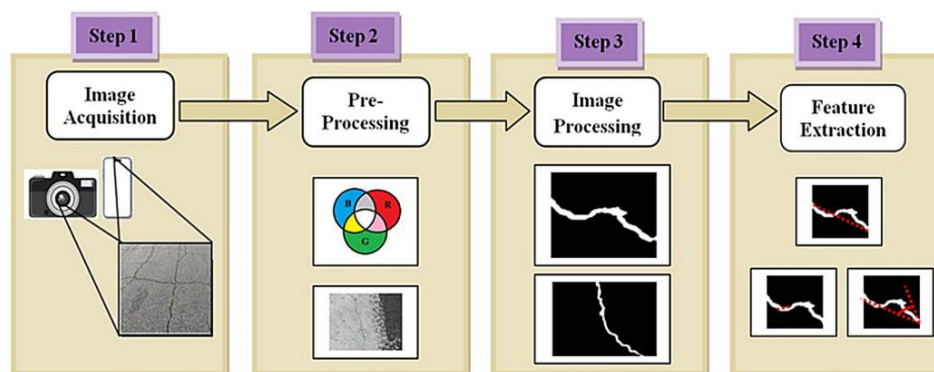
Nota. Tomado de "A review study of intelligent road crack detection: Algorithms and systems" por E. Hemdan. 2025.

En esta investigación se presenta como complemento en la Figura 2 el desarrollo de la transformación y predicción de imágenes, en donde se inicia desde la recopilación de

imágenes, se hace el procesamiento de ellas y se extraen las propiedades relevantes para su clasificación. (El-Din Hemdan & Al-Atroush, 2025)

Figura 2.

Esquema de desarrollo de detección de grietas



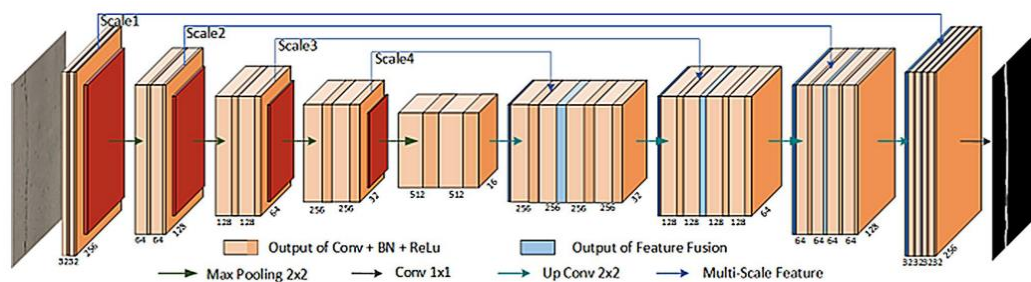
Nota. Tomado de "A review study of intelligent road crack detection: Algorithms and systems" por E. Hemdan. 2025.

Esto es posible gracias al uso de la red convolucional U-net, donde el objetivo es lograr un análisis eficiente de imágenes y reconstrucción por medio de capas Convolucionales y de Batch Normalization según se presenta en la Figura 3. (El-Din Hemdan & Al-Atroush, 2025)

Puntualmente se busca la reducción de dimensionalidad de la imagen, para luego recuperarla en el formato de contraste, para facilidad de detección.

Figura 3.

Estructura de U-Net



Nota. Tomado de "A review study of intelligent road crack detection: Algorithms and systems" por E. Hemdan. 2025.

Para la captura de la información se requiere de la recolección de fotos del pavimento por medio de cámaras u otros elementos montados en Vehículos o drones. (Ramírez-Villanueva et al., 2025) Luego, se procesa la información con métodos de normalización, reducción de ruido, conversión de formato, entre otras para finalizar con la detección de las grietas utilizando sistemas de Aprendizaje de Máquina y Deep Learning, basado en la categorización de las grietas y revisión de severidad con altos estándares y desempeño. (El-Din Hemdan & Al-Atroush, 2025).

Se identifica en este estudio que el uso de estas tecnologías es clave ya que en las construcciones civiles es importante garantizar la seguridad y durabilidad de la infraestructura, donde el principal reto es la obtención de la información para entrenar los modelos enfocados en las grietas superficiales. Si se requiere el análisis otro tipo de grietas como las que están por debajo del asfalto se requiere el uso de radares como el Ground Penetrating Radar (GPR) para obtener la información. No obstante, el uso de herramientas de Deep Learning y GANs se presentan como una alternativa para tener una mayor cantidad de imágenes a partir de datos sintéticos cercanos a los reales. (El-Din Hemdan & Al-Atroush, 2025).

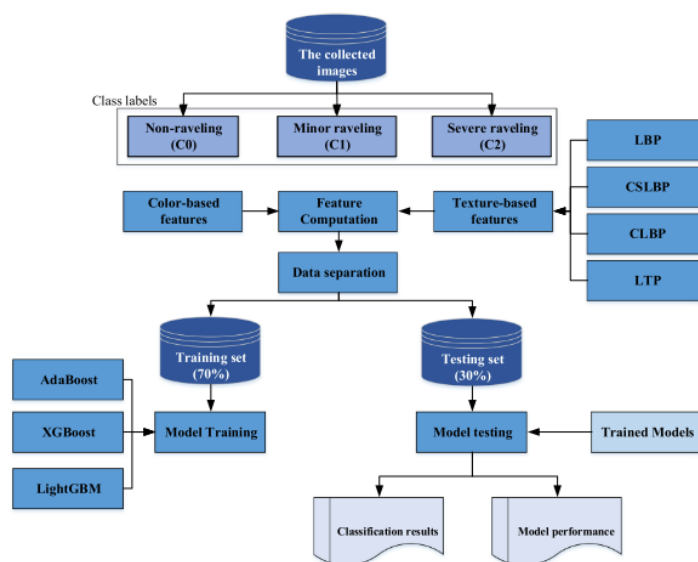
Por otro lado, en (H. Nhat-Duc & Van-Duc, 2023) se hace la revisión del uso de modelos para la detección de pavimento deteriorado basado en visión de computadora y de métodos para mejorar el desempeño de los modelos de clasificación, como lo es la transformación de las imágenes para que los modelos tenga una mejor capacidad de predecir las fallas utilizando los modelos de la familia Gradient Boosting como XGBoost. (H. Nhat-Duc & Van-Duc, 2023; Hoang & Tran, 2022).

En la Figura 4 se muestra el flujo de la construcción del modelo de (H. Nhat-Duc & Van-Duc, 2023), en donde se presenta el tratamiento de la información de las imágenes para poder utilizar modelos diferentes a Redes Neuronales como XGBoost en conjunto con un framework de división del set de entrenamiento y prueba para identificar el desempeño de la clasificación,

tomando como base que no se detecta la ubicación de las grietas, sino que si hay al menos una en las imágenes de insumo.

Figura 4.

Desarrollo de análisis de grietas con modelos XGBoost



Nota. Tomado de "Computer Vision-Based Severity Classification of Asphalt Pavement Raveling Using Advanced Gradient Boosting Machines and Lightweight Texture Descriptors" por H. Nhat-Duc. 2023.

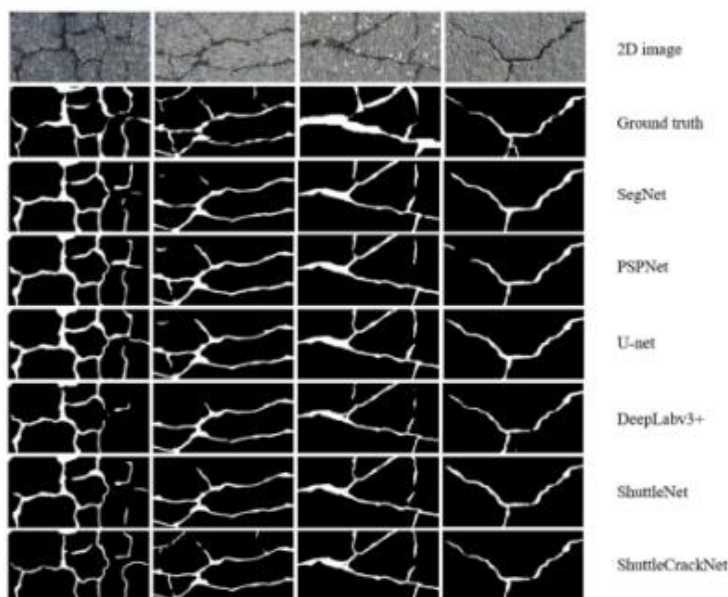
Como punto base, se han utilizado métodos asociados con redes neuronales convolucionales y modelo ADABOOST, con la limitante que se requieren datasets especializados para tener un comportamiento adecuado, de tres dimensiones. Pero con el tratamiento de los datos de entrada se logran obtener predicciones con alta precisión mientras se mantiene un costo computacional sin recurrir a modelos de redes neuronales.

Adicional, se presentan casos en los que la vista de estas grietas está sujeta a las variables meteorológicas de cuando se toma la información, por ello se han innovado en sistemas como el presentado en (Zhang et al., 2025), donde se crea un modelo que utiliza mecanismos de atención como los utilizados en modelos de Lenguaje Natural para reducir el nivel de ruido propio de los dataset y mejorar la representabilidad de las grietas como se

muestra en la Figura 5. Este problema se presenta dentro de la línea de innovación en la inteligencia artificial y el Deep Learning para la detección de grietas a nivel de píxeles con el reto de la diversidad y complejidad de las vías a analizar en conjunto con el desbalance de grietas sobre las imágenes que no tienen. (Fakhri et al., 2025).

Figura 5.

Ejemplo de identificación de grietas por diferentes modelos



Nota. Tomado de “Pixel-Level Intelligent Recognition of Asphalt Pavement Cracks with an Improved ShuttleNet” por A. Zhang. 2025.

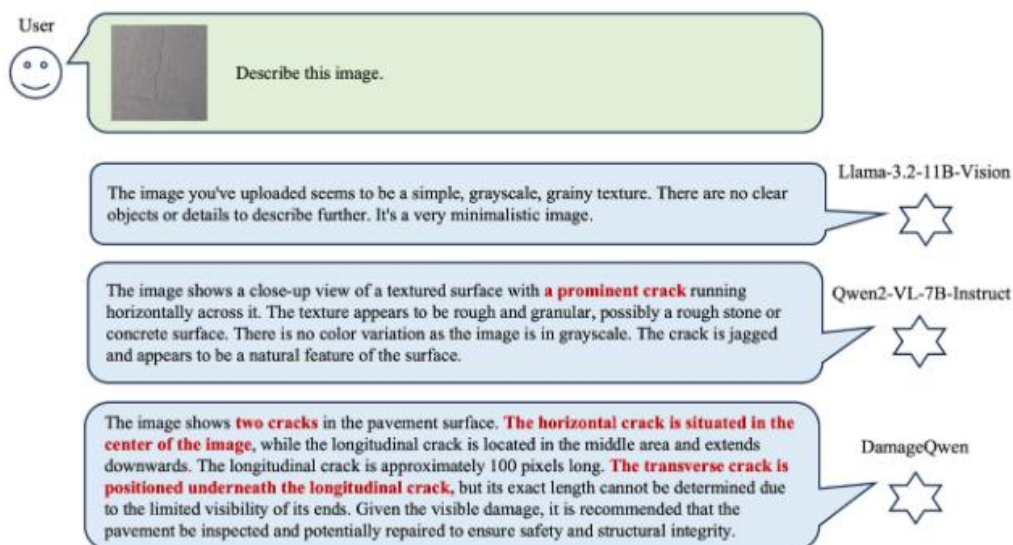
Puntualmente, se menciona la limitante de los métodos tradicionales de Redes Neuronales Convencionales que se enfocan en sólo algunas partes de las imágenes mientras que los nuevos métodos van hacia el uso de mecanismos de codificación – decodificación para identificar la mayor información posible, si bien se presenta un reto de optimización de recursos ya que en el momento es una alternativa costosa computacionalmente y que, de acuerdo con los data set con que se entrene, tiende a tener un desempeño igual o peor que otros modelos tradicionales de Redes Neuronales.

Como alternativa, se han desarrollado métodos para reducir el impacto que tiene el etiquetado de la información de grietas y de la interacción con los usuarios para tomar las mejores decisiones. Por ejemplo, en (Zhang & Liu, 2025) se diseñó el modelo DamageQwen donde se busca un modelo de Visión de computadora basado en lenguaje natural. Esta última necesidad surgió porque la inspección periódica de las vías conlleva una larga cantidad de recursos humanos y tecnológicos, contando las variables asociadas con la opinión experta que puede llevar a diagnósticos incorrectos como se indicó al inicio de la sección.

En ésta se combinan los modelos asociados con la detección de grietas de distintos tipos como el YOLO v11x y que traduce el contenido a Lenguaje Natural para que lo pueda analizar una persona con distintos niveles de experiencia. En la Figura 3 se muestra un ejemplo:

Figura 6.

Ejemplo de análisis de imágenes de grietas

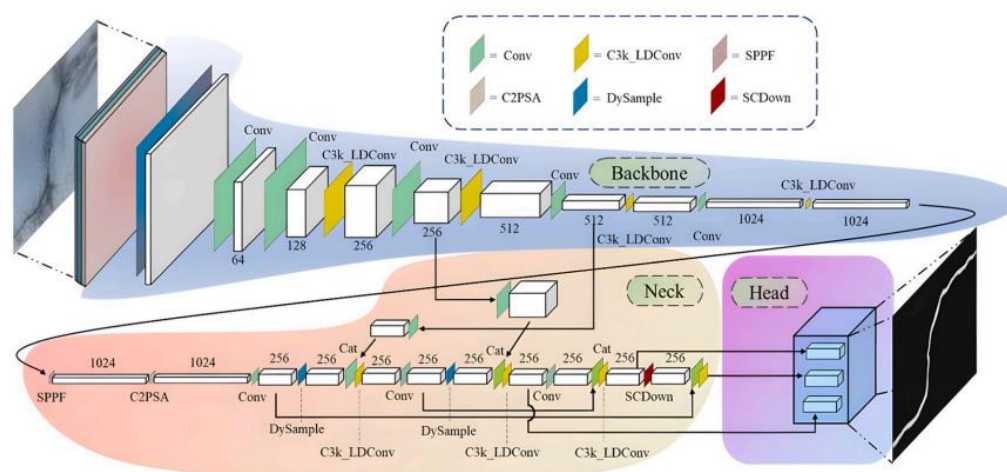


Nota. Tomado de "Vision-enhanced multi-modal learning framework for non-destructive pavement damage detection" por Y. Zhang. 2025.

Con respecto a los modelos de visión de computadora, en la Figura 7 se presenta una variación del modelo de YOLOv11 llamado YOLO-LY, que está diseñado como un modelo de segmentación o arquitectura unificada que eficientemente genera cajas de detección y segmentación a nivel de píxeles mientras disminuye los parámetros del modelo y requerimientos computacionales sin comprometer el desempeño de éste. (Yin et al., 2026)

Figura 7.

Topología de red Yolo-LY



Nota. Tomado de "A UAV-deployable lightweight framework for real-time bridge crack detection via YOLO-LY algorithm" por X. Yin. 2026.

Puntualmente esta investigación permitió crear un nuevo paradigma de mantenimiento de forma automática, balanceando procesamiento en tiempo real y operación segura. Por ejemplo, en un caso de uso en donde se hace la inspección de grietas en puentes de difícil acceso hubo una reducción del costo total del 94.5% utilizando herramientas de Visión de Computadora, mejoró la eficiencia de la inspección en un 85% y redujo los errores en un 5.7% basado en el personal, tiempo y materiales requeridos para hacer la inspección. (Yin et al., 2026)

Marco normativo e institucional

La presente investigación se desarrolla en el marco institucional de la gestión y conservación de la infraestructura vial urbana de Bogotá D.C., cuya entidad competente es el Instituto de Desarrollo Urbano (IDU). Esta entidad define los lineamientos técnicos, metodológicos y contractuales para el diagnóstico, evaluación y determinación del estado de los pavimentos que conforman la malla vial y el espacio público de la ciudad.

En la Guía GU-IC-15 “Metodología de diagnóstico y determinación del estado de los pavimentos y otros elementos del espacio público de Bogotá D.C.”, el IDU establece como referente normativo para el diagnóstico superficial de pavimentos flexibles y rígidos la norma ASTM-D-6433. (IDU, 2021)

Esta referencia se ratifica en la sección de referencias bibliográficas del documento institucional, donde se incluye formalmente la norma *ASTM D 6433*.

Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys

La adopción de la norma ASTM D6433 implica que el diagnóstico superficial oficial debe realizarse conforme a la metodología del Índice de Condición del Pavimento (PCI), la cual establece la identificación tipológica de deterioros, su clasificación por niveles de severidad y el cálculo de valores de deducción que permiten obtener un índice numérico en escala de 0 a 100. Este índice constituye el parámetro técnico que soporta la toma de decisiones de mantenimiento, rehabilitación o reconstrucción. (ASTM, 2018)

En el Apéndice I del documento contractual, el IDU define los rangos de clasificación por condición superficial con base en el valor del PCI, estableciendo categorías desde “Rojo” (PCI 0–25) hasta “Verde” (PCI 86–100), las cuales determinan la priorización de intervenciones y la necesidad de estudios estructurales complementarios. En consecuencia, el resultado técnicamente vinculante dentro del esquema institucional no es únicamente la detección de

fisuras, sino la correcta estimación del PCI conforme al procedimiento normativo adoptado. (IDU, 2025)

Adicionalmente, el marco contractual exige la realización de actividades de calibración y patronamiento de equipos previo a la ejecución del diagnóstico, conforme a la Guía GU-IC-13, estableciendo que solo se aceptará información proveniente de equipos que hayan participado en dicho proceso. Esta disposición introduce un requisito de control técnico y validación previa que garantiza la confiabilidad de los datos empleados en la toma de decisiones que resulten del diagnóstico contratado.

No obstante, aunque el marco institucional define con claridad la metodología tradicional de inspección visual bajo la norma ASTM D6433, no se identifican lineamientos específicos que regulen la incorporación de sistemas basados en inteligencia artificial para el reconocimiento automático de deterioros superficiales. Las disposiciones vigentes están estructuradas bajo el supuesto de una evaluación realizada por inspectores capacitados y no contemplan criterios explícitos para la validación de modelos algorítmicos, la trazabilidad de datos de entrenamiento, ni la equivalencia metodológica entre segmentación digital y clasificación de deterioros por severidad. (IDU, 2025)

Este vacío normativo genera una brecha entre el desarrollo tecnológico y el marco regulatorio aplicable. Mientras que los modelos de inteligencia artificial son evaluados mediante métricas computacionales como precisión, F-measure o IoU, el esquema institucional continúa fundamentando sus decisiones en el valor del PCI y su clasificación correspondiente.

En consecuencia, la implementación de tecnologías automatizadas exige demostrar no solo desempeño algorítmico, sino compatibilidad técnica con la metodología oficial adoptada por el IDU.

En este contexto, la presente investigación se sitúa en la intersección entre el marco normativo vigente y la transformación tecnológica del proceso de diagnóstico, examinando, desde la evidencia disponible, el desempeño de los modelos basados en deep learning en

relación con los criterios de evaluación empleados en el procedimiento ASTM D6433 para la clasificación de grietas, adoptado institucionalmente. De esta manera, el estudio analiza la precisión del reconocimiento automático de grietas y contribuye a la discusión académica sobre los desafíos técnicos asociados a la adopción de tecnologías automatizadas en el diagnóstico de daños superficiales en pavimentos flexibles.

Marco teórico

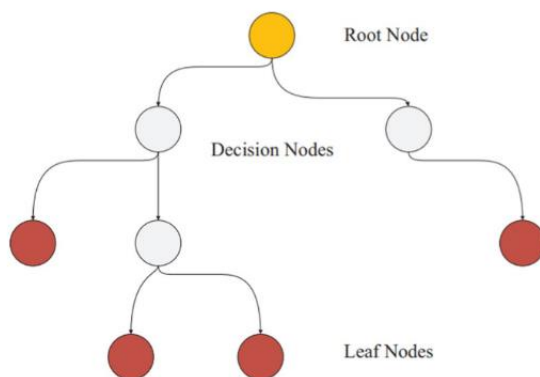
El aprendizaje de máquina es utilizado para la evaluación de información y de identificación de patrones que puedan servir para la toma de decisiones, siendo útil para campos como el de la detección de grietas en pavimentos ya que pueden aprender continuamente de la información de entrada. Puntualmente, se divide en el campo de supervisado y no supervisado, siendo el primero correspondiente de predecir salidas a partir de información etiquetada o conjuntos de entrada – salida definidos mientras que la segunda consiste en utilizar la información de entrada para identificar patrones como lo es la creación de clústeres (Zeida et al., 2024).

Puntualmente para la revisión de grietas en pavimentos flexibles se han utilizado diferentes algoritmos para hacer la clasificación de imágenes, que se resumen a continuación: (Zeida et al., 2024)

Árboles de decisión: Corresponde a un modelo con estructura de árbol construido a partir de raíces, hojas y ramas. En ellos para poder predecir un valor se sigue un camino basado en evaluación de condiciones hasta llegar a un valor final. Un ejemplo se presenta en la Figura 9.

Figura 9.

Esquema de Árbol de decisión



Nota. Tomado de "Machine learning modeling of transverse cracking in flexible pavement" por Zeida 2024

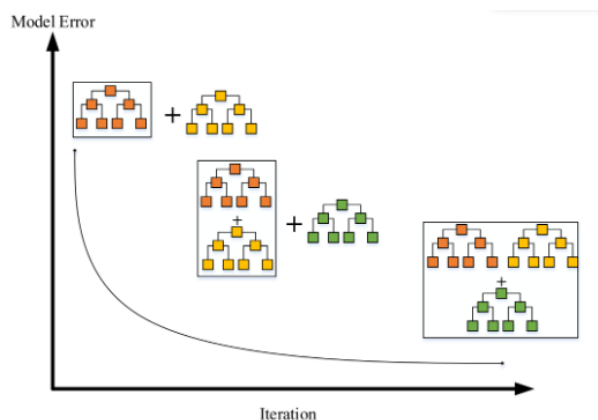
Support Vector Machine (SVM): Las máquinas de vector de soporte utilizan el concepto de minimización de riesgo estructural para lograr la identificación de patrones y generalización por medio de patrones de aprendizaje. Matemáticamente consiste en calcular una función de regresión lineal en un espacio con varias características o dimensiones, con la información de entrada representada por una función no lineal.

Ensamblados: Una técnica de ensamble corresponde a la unión de modelos débiles donde al unir sus estimaciones se obtiene una predicción más precisa, donde el método tradicional es de promediar sus resultados cuando se trata de calcular una variable numéricas. Dentro de estos ensambles se han creado otros tipos de modelo como de potenciación del gradiente (Gradient Boost por su nombre en inglés)

Extreme gradient boost machine (XGBoost) Este algoritmo genera un conjunto de árboles de clasificación de forma iterativa y combina sus predicciones de forma aditiva para obtener la principal etiqueta como se presenta en la Figura 10. Gracias a la capacidad de trabajar con clasificación multivariante y no lineal ha logrado tener desempeño alto en trabajos complejos de modelación de datos. (El-Din Hemdan & Al-Atroush, 2025)

Figura 10.

Muestra del proceso de entrenamiento del XGBoost

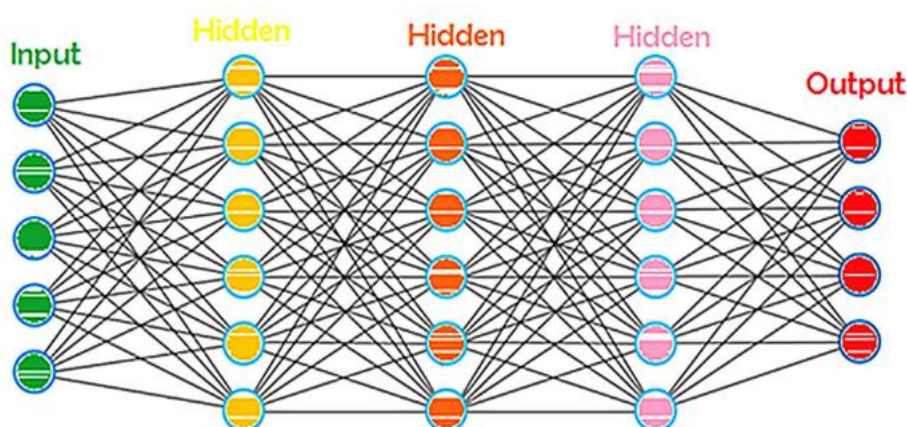


Nota. Tomado de "A review study of intelligent road crack detection: Algorithms and systems" por El-Din Hemdan. 2025

Como complemento al Machine Learning tradicional se ha desarrollado la rama del Deep Learning, donde se busca emular la capacidad del cerebro humano de procesar entradas complejas como imágenes en clasificación e identificación. Esto se logra a partir de la construcción de arquitecturas como la que se presenta en la Figura 11, donde se presentan los componentes de estas redes: (Hoang & Tran, 2025)

Figura 11.

Arquitectura Red Neuronal



Nota. Tomado de “Computer vision based asphalt pavement segregation detection using image texture analysis integrated with extreme gradient boosting machine and deep convolutional neural networks” por Hoang N. 2025

A su vez, estas se dividen en categorías:

Artificial Neural Networks (ANN): Las redes Neuronales Artificiales son el caso general, donde tienen la capacidad de procesar información de forma paralela, permitiendo que sea resiliente a fallas mientras aprende de información histórica para toma de decisiones.

Convolutional Neural Networks (CNN): Las redes Convolucionales están especializadas para el análisis de imágenes, extrayendo información como ejes, texturas y formas. A partir de su arquitectura logran una alta precisión, son robustas contra el ruido y se pueden reentrenar bajo el principio de Transfer Learning.

Generative Adversarial Networks (GAN): Las redes generativas antagónicas utilizan un generador y discriminador para que compitan entre ellos para generar información realista, es decir, son una fuente para la generación de datos sintéticos por medio del análisis de la distribución de estos.

Recurrent Neural Networks (RNN): Las Redes Neuronales Recurrentes están diseñadas específicamente para el análisis de datos secuenciales, logrando recordar información histórica de distinta extensión, aunque tiene retos para recordar dependencias a largo plazo.

Adicional, para la evaluación de estos modelos se utiliza en primera instancia la precisión (Accuracy) en donde se mide la proporción de las predicciones realizadas de forma correcta (TP como verdadero positivo y TN como verdadero negativo) sobre el total de predicciones (Morshed & Bheemasetti, 2025):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Como complemento a esta se utiliza otra forma de calcular la precisión (Precision) donde se analiza la capacidad del modelo de predecir los Positivos de forma adecuada, o en este caso de las Grietas en el asfalto:

$$Precision = \frac{TP}{TP + FP}$$

El recall de forma diferente identifica la capacidad de detectar solamente los positivos, siendo clave para no dejar sin considerar ninguna condición crítica.

$$Recall = \frac{TP}{TP + FN}$$

Finalmente, en el campo de indicadores se utiliza el F1- Score donde se calcula la media armónica entre el Recall y Precision, basado en medir el impacto de falsos positivos y negativos:

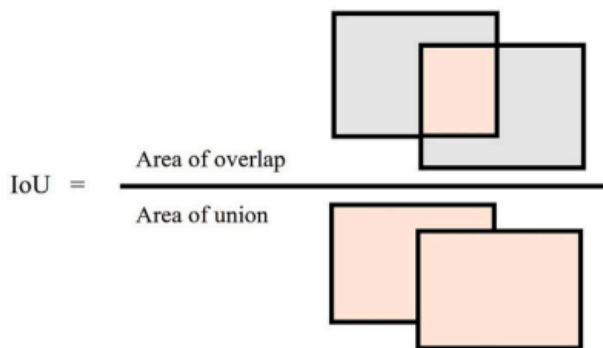
$$F1\ score = \frac{Precision \cdot Recall}{Precision + Recall}$$

En el campo de la visión por computadora el modelo de YOLO (You Only Look Once por sus siglas en inglés) ha transformado la detección de objetos, mejorando la predicción en escenarios en tiempo real. Puntualmente, se introdujo la primera versión de YOLO en 2016, transformando el paradigma de detección de objetos a un problema de regresión por medio de la división de imágenes en una malla de tamaño $S \times S$, en conjunto con la predicción de categorías, generación de coordenadas de ubicación de los objetos y puntajes de confianza (Confidence Score) (Tian et al., 2026).

Para la evaluación de estos modelos, en conjunto con las métricas tradicionales de Accuracy, Precision, Recall y F1-Score, se utilizan otras tales como la intersección sobre unión IoU que corresponde al ratio entre el área de solapado de dos cajas de frontera sobre el área total que cubren, mostrado de forma gráfica en la Figura 12. El objetivo es que esta métrica sea lo más alta posible ya que corresponde a que la localización predicha por el algoritmo es cercana a la real.

Figura 12.

Esquema de intersección sobre Unión



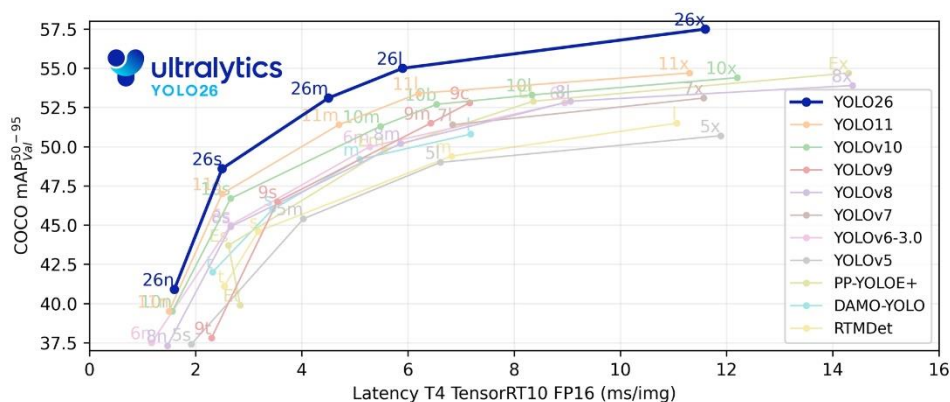
Nota. Tomado de "*Ultralytics*" de Jocher, G. 2026

La otra métrica utilizada para el análisis es AP, que corresponde al área bajo la curva de las métricas de Precisión y Recall sumado a la mAP como el promedio de desempeño considerando todas las posibles categorías. Por ello, esta última se utiliza frecuentemente para

la evaluación de los modelos de YOLO (Tian et al., 2026). En la Figura 13 se muestra el desempeño de los principales modelos desarrollados por el proyecto de ultralytics, utilizando como base el dataset de COCO (Common Objects in Context).

Figura 13.

Desempeño de modelos de referencia YOLO



Nota. Tomado de "Ultralytics" de Jocher, G. 2026

A su vez, se mide el desempeño del modelo de acuerdo con el tiempo de predicción de objetos para cada una de las imágenes o frames en el caso que se analice un video, aspecto que varía de acuerdo con el tamaño de dichos modelos, es decir, cantidad de parámetros y tamaño de la red neuronal. Por ejemplo, en el caso de la versión más reciente del modelo de YOLO publicada en enero de 2026, los modelos varían de talla desde a n (nano) hasta la x(grande). Este modelo como principal tiene las siguientes ventajas, que influirán en la aplicación en la detección de grietas en pavimentos flexibles (Jocher & Qiu, 2026):

- **Simplicidad:** Eliminación de pasos previos de tratamiento de las imágenes para que la inferencia sea ligera, más rápida y fácil de implementar en sistemas reales
- **Facilidad de implementación:** basado en la simplificación del pipeline para entrenamiento y fine-tuning de modelos

- Optimización de acuerdo con actividades específicas como Segmentación y OBB (Oriented Bounding Boxes)

Metodología

En primera instancia se presentan los puntos clave a considerar en el desarrollo del proyecto:

Enfoque metodológico

La investigación se desarrolla bajo un enfoque cuantitativo, ya que busca comparar la precisión en la detección de grietas entre un modelo de visión por computadora basado en Deep Learning y la inspección visual tradicional. Para ello, se utilizan métricas cuantificables que permiten evaluar de manera objetiva el desempeño de ambos métodos sobre un mismo conjunto de imágenes.

Tipo y diseño de la investigación

El estudio es de tipo descriptivo y comparativo, debido a que se analizan las características del desempeño de cada método y luego se comparan sus resultados en términos de precisión.

El diseño es no experimental, ya que no se manipulan las variables, sino que se trabaja con resultados obtenidos a partir de la aplicación del modelo sobre imágenes reales. Además, es de corte transversal, dado que el análisis se realiza en un único momento sobre un conjunto específico de imágenes.

Supuestos de la investigación

Para facilitar el desarrollo del análisis, se plantean los siguientes supuestos:

- Las imágenes utilizadas representan condiciones típicas de pavimentos flexibles.
- La calidad de las imágenes (resolución e iluminación) es adecuada para la detección de grietas.
- El modelo de Deep Learning de detección de grietas mantiene un comportamiento estable durante su ejecución.
- La inspección visual tradicional se toma como referencia para identificar grietas.

- Se asume que las referencias de grietas obtenidas mediante inspección visual tradicional representan de forma adecuada la presencia real de grietas.

Estos supuestos permiten simplificar el análisis sin afectar de manera significativa la comparación entre los métodos.

Variables técnicas y contextuales

Las variables técnicas y contextuales que se van a considerar en este trabajo para comparar la precisión entre el modelo de Deep Learning y la inspección visual tradicional se presentan en la Tabla 1 y Tabla 2, respectivamente.

Por un lado, las variables técnicas están relacionadas con el modelo que se va a utilizar, en este caso Yolov26, incluyendo aspectos como su funcionamiento, su entrenamiento, métricas del modelo y las condiciones de las imágenes de entrada. Por otro lado, las variables contextuales tienen más que ver con el entorno en el que se presentan las grietas, como el tipo de pavimento, las condiciones en las que se capturan las imágenes y también el criterio del evaluador, que al final introduce un componente humano que puede influir en la interpretación de los resultados.

Tabla 1.

Descripción de variables técnicas que afectan el análisis de grietas en pavimento flexible

Variable	Descripción	Papel en el problema
Tipo de modelo de detección	Hace alusión al método o algoritmo que se utiliza para identificar las grietas en las imágenes del pavimento. Pueden ser métodos más tradicionales de procesamiento de imágenes o modelos más recientes basados en Deep learning.	Dependiendo del modelo que se utilice, el sistema puede detectar más o menos grietas, especialmente cuando son muy pequeñas o poco visibles.
Entrenamiento del modelo	Tiene que ver con cómo se entrenó el modelo, por ejemplo, cuántas imágenes se usaron y qué tan variadas eran.	Si el modelo fue entrenado con buenos datos y suficientes ejemplos, normalmente va a detectar mejor las grietas en nuevas imágenes.
Parámetros del modelo	Son configuraciones del modelo, como qué tan sensible es al detectar grietas o el tamaño de imagen que utiliza para analizar.	Ajustar estos parámetros puede hacer que el modelo detecte más grietas o que sea más estricto al momento de identificarlas.

Variable	Descripción	Papel en el problema
Calidad y preparación de la imagen	Incluye el nivel de detalle de la imagen (resolución) y los procesos previos como mejora de contraste o reducción de ruido.	Una imagen con buena resolución y bien procesada permite que las grietas se vean más claras, mientras que una mala calidad puede hacer que algunas no se detecten.
Métricas de evaluación	Son las medidas que se utilizan para evaluar qué tan bien funciona el modelo, por ejemplo, precisión o recall.	Permiten comparar los resultados entre diferentes métodos de detección y ver cuál funciona mejor.

Nota: Elaboración propia

Tabla 2.

Descripción de variables contextuales que afectan el análisis de grietas en pavimento flexible

Variable	Descripción	Papel en el problema
Tipo de pavimento	Se refiere a las características de la superficie del pavimento, como su textura o nivel de desgaste.	Dependiendo del tipo de pavimento, las grietas pueden ser más fáciles o difíciles de identificar en las imágenes.
Condiciones visuales de la superficie	Incluye tanto la iluminación (sombras, poca luz) como la presencia de elementos en el pavimento como suciedad, manchas o agregados del asfalto.	Estas condiciones pueden dificultar que las grietas se vean claramente o hacer que el sistema confunda otros elementos con grietas, afectando la precisión de la detección.
Tipo de grieta	Forma o patrón de la grieta, por ejemplo, longitudinal, transversal o en forma de red.	Algunos tipos de grietas son más fáciles de detectar que otros dependiendo del método utilizado.
Captura de las imágenes	Forma en que se toman las fotografías del pavimento, por ejemplo, el ángulo, la distancia o si la imagen está bien enfocada.	Si las imágenes no se toman bien, puede ser más difícil detectar las grietas correctamente.
Criterio del evaluador	Interpretación que hace la persona que revisa o valida los resultados del modelo.	En algunos casos, diferentes personas pueden interpretar de forma distinta si una grieta fue detectada correctamente o no.

Nota: Elaboración propia

Método de recolección de información

Específicamente, la recolección de imágenes se basa en el uso de cámaras de alta resolución (tipo GoPro) montadas sobre vehículos en movimiento, con registro GPS incorporado, siguiendo los estándares establecidos en la norma ASTM D6433-20 para la evaluación superficial de pavimentos flexibles, esto debido a que un modelo de Deep Learning para el reconocimiento de patologías en pavimentos requiere de datos con características técnicas específicas en su proceso de captura, que se resumen en la Tabla 3:

Tabla 3.*Resumen de características imágenes de predicción de grietas*

CARACTERÍSTICA	IMÁGENES DE PREDICCIÓN
Propósito	Ser procesadas por el modelo para generar diagnósticos
Requisito de etiquetado	No – procesadas automáticamente
Diversidad requerida	Condiciones reales del corredor evaluado
Consistencia técnica	Estandarizada
Volumen	Variable – según longitud del corredor a evaluar

Nota: Elaboración propia

Descripción del equipo de captura

El equipo de captura empleado consiste en una o varias cámaras de acción tipo GoPro (modelo Hero 9 Black o superior), montadas sobre un vehículo de inspección. Este sistema permite registrar video continuo de alta resolución (4K o 2.7K) a una tasa de entre 60 y 120 fotogramas por segundo, con GPS incorporado para la geolocalización automática de cada cuadro. Las características principales del sistema son:

- **Resolución de video:** mínimo 2.7K (2704 x 1520 píxeles), preferiblemente 4K (3840 x 2160 píxeles).
- **Velocidad de captura:** entre 60 fps (fotogramas por segundo) y 120 fps para garantizar cobertura continua del pavimento a velocidades de recorrido de 10 a 30 km/h.
- **Campo visual (FOV):** modo Wide o Linear, para minimizar la distorsión periférica en el análisis posterior.
- **GPS integrado:** registro de coordenadas geográficas (latitud, longitud) en cada fotograma, con precisión de ± 3 metros.

- **Almacenamiento:** tarjeta microSD de alta velocidad (clase U3, V30 o superior) de al menos 256 GB.
- **Fuente de alimentación:** batería externa o conexión al sistema eléctrico del vehículo para recorridos de larga duración.

Configuración de montaje

La cámara se instala en el vehículo de inspección mediante un soporte rígido ajustable ubicado en la parte delantera o trasera, orientado hacia la calzada. Los parámetros de montaje son:

- **Altura de montaje:** entre 1.0 m y 1.5 m sobre el nivel del pavimento, para garantizar una perspectiva cenital adecuada.
- **Ángulo de inclinación:** 15° a 25° respecto a la vertical, orientado hacia el frente de avance del vehículo, de modo que se capture el carril completo dentro del encuadre.
- **Cobertura lateral:** la cámara debe capturar el ancho total del carril de inspección (mínimo 3.0 m a 3.25 m según PMT), o ambos carriles cuando la calzada lo permita.
- **Estabilización:** uso de estabilizador de imagen electrónico (EIS) activado en la cámara, complementado con soporte amortiguado para reducir la vibración producida por imperfecciones en el pavimento.

Objetivo de la captura para entrenamiento

El conjunto de imágenes de entrenamiento constituye el insumo fundamental para el desarrollo del modelo de Deep Learning. Su objetivo es proveer al algoritmo de ejemplos representativos y suficientemente variados de todas las patologías que el modelo deberá reconocer para las fisuras o grietas.

Parámetros técnicos de captura para entrenamiento

Para garantizar la calidad y homogeneidad del dataset de entrenamiento, la captura debe realizarse bajo los siguientes parámetros:

- **Condiciones ambientales y lumínicas**

1. **Iluminación:** luz natural diurna, evitando capturas entre las 11:00 a.m. y las 2:00 p.m. para reducir el efecto de sombras duras y reflejos especulares sobre el asfalto.
2. **Condición del pavimento:** el pavimento debe estar seco. Se evita la captura en condiciones de lluvia o con pavimento mojado, ya que el agua modifica la apariencia de las patologías y puede inducir falsos positivos en el modelo.
3. **Cielo:** se prefieren condiciones de cielo nublado difuso, que proveen iluminación homogénea sin sombras definidas.
4. **Temperatura:** la captura se realiza preferiblemente en las horas de la mañana (7:00 a.m. a 10:00 a.m.) o en la tarde (3:00 p.m. a 5:00 p.m.) para evitar la dilatación térmica que puede generar brillos sobre el asfalto.

- **Parámetros de velocidad y frecuencia de captura**

1. **Velocidad del vehículo:** entre 10 y 20 km/h para maximizar la resolución efectiva por metro lineal de pavimento.
2. **Frecuencia de extracción de fotogramas:** 1 fotograma por cada 20 a 30 cm de avance longitudinal, equivalente a una extracción de 1 cuadro cada 5 a 8 fotogramas del video original a 60 fps con velocidad de 15 km/h.
3. **Traslape entre imágenes consecutivas:** mínimo 30% del área longitudinal entre un cuadro y el siguiente, para garantizar cobertura continua sin zonas ciegas.

4. **Resolución final de las imágenes extraídas:** mínimo 1280 x 720 píxeles (HD), preferiblemente 1920 x 1080 píxeles (Full HD).

Proceso de etiquetado para entrenamiento

Una vez extraídas las imágenes, se realiza el proceso de etiquetado (anotación) manual por parte de especialistas en pavimentos, siguiendo los criterios de la norma ASTM D6433-20:

- **Herramienta de etiquetado:** se recomienda el uso de plataformas especializadas como Labellmg, Roboflow, CVAT o Label Studio, que permiten generar anotaciones en formato YOLO, COCO o Pascal VOC según la arquitectura del modelo seleccionado que en nuestro caso se usara YOLO.
- **Tipo de anotación:** para modelos de detección de objetos (bounding boxes), se dibuja un rectángulo delimitador sobre cada instancia de patología identificada, indicando la clase (tipo de daño).
- Para modelos de segmentación semántica, se realiza la delineación del área afectada píxel a píxel, lo que permite calcular directamente el porcentaje de área afectada por tipo de patología.
- **Protocolo de etiquetado:** mínimo dos especialistas revisan de forma independiente cada imagen, y se aplica consenso o revisión por un tercer experto en caso de discrepancia.
- **Volumen mínimo recomendado:** al menos 500 para el modelo de detección.

Procedimiento de captura para predicción

El procedimiento específico para la captura de imágenes destinadas a la predicción es el siguiente:

1. **Verificación del tramo a evaluar:** se confirma la localización GPS del segmento vial (PK de inicio y fin, CIV, nombre de la vía), el ancho de calzada, el número de carriles y el tipo de pavimento.
2. **Configuración del equipo:** se aplican exactamente los mismos ajustes de la cámara utilizados durante la captura de entrenamiento. Se verifica el nivel de batería o la conexión de alimentación, el espacio disponible en la tarjeta de memoria, y el estado del GPS (señal activa confirmada).
3. **Recorrido de inspección:** el vehículo realiza al menos una pasada completa por cada carril del segmento, a velocidad constante de 10 a 20 km/h. Para corredores con dos o más carriles, se realiza una pasada independiente por cada carril.
4. **Registro de metadatos:** el operador registra en planilla de campo la fecha, hora de inicio y fin del recorrido, condiciones ambientales (temperatura, nubosidad, estado de la superficie), y cualquier incidencia que pueda afectar la calidad de las imágenes (obstáculos, vehículos estacionados, zonas con agua o suciedad excesiva).
5. **Procesamiento por el modelo de Deep Learning:** las imágenes son procesadas por el modelo entrenado, generando como salida la lista de deterioros identificados con sus respectivas dimensiones en ml, el tipo de patología, el nivel de severidad y las coordenadas GPS de localización.

Control de calidad de las imágenes de predicción

Antes de procesar las imágenes con el modelo, se realiza una revisión de calidad que verifica:

- **Ausencia de desenfoque de movimiento:** imágenes nítidas sin rastros de movimiento. Las imágenes con desenfoque pronunciado se descartan.

- **Cobertura completa del carril:** se confirma que el 100% de la longitud del segmento fue capturada, sin zonas sin cobertura mayores a 1 metro lineal.
- **Exposición adecuada:** se descartan imágenes sobreexpuestas (detalle perdido en zonas claras) o subexpuestas (detalle perdido en zonas oscuras), reemplazándolas por fotogramas adyacentes del video original.
- **Ausencia de oclusiones:** se identifican y gestionan las imágenes con oclusiones por vehículos, peatones, sombras de estructuras o elementos ajenos al pavimento que cubran más del 20% del área de la imagen.
- **Consistencia GPS:** se verifica que la secuencia de coordenadas GPS sea continua y coherente con la longitud y dirección del segmento evaluado.

Localización de las imágenes capturadas

Localización general — Bogotá D.C.

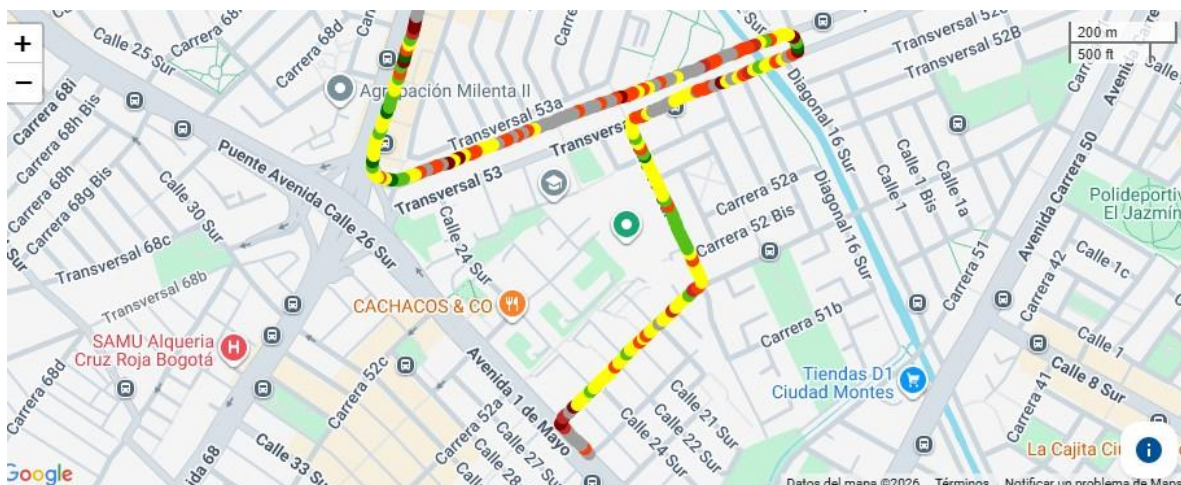
La ciudad de Bogotá D.C. constituye la capital de Colombia y el principal centro urbano del país, con una extensión aproximada de 1.775 km² y una población superior a los 8 millones de habitantes. Políticamente se divide en 20 localidades, cada una con características propias de densidad poblacional, uso del suelo y condición de la infraestructura vial. La malla vial del Distrito Capital supera los 15.000 km de vías distribuidos entre arterias, colectoras y locales, cuya gestión, diagnóstico y mantenimiento están a cargo del Instituto de Desarrollo Urbano (IDU).

La campaña de captura de imágenes para el entrenamiento y la predicción del modelo YOLOv26 se realizó en el suroriente del Distrito Capital, en el sector comprendido entre la localidad de Kennedy y la localidad de Puente Aranda, sobre corredores que hacen parte de la malla vial local y colectora del área. La Figura 15 presenta la localización general del conjunto

de tramos evaluados, con la codificación cromática de condición superficial según el Índice de Condición del Pavimento (PCI) definido en la norma ASTM D6433-20.

Figura 14.

Localización general — Tramos evaluados sobre Av. 1 de Mayo y Transversal 53, Bogotá D.C.



Nota. Fuente: elaboración propia con base en Google Maps © 2026.

Los corredores evaluados se ubican en el área de influencia de dos localidades contiguas: Kennedy y Puente Aranda. La localidad de Kennedy, situada en el suroccidente de Bogotá, es la localidad más poblada del Distrito con aproximadamente 1,2 millones de habitantes. Su territorio se caracteriza por un uso del suelo predominantemente residencial de estratos 2 y 3, con importante presencia de actividad comercial e industrial de pequeña y mediana escala. La malla vial local de esta localidad presenta uno de los mayores índices de deterioro superficial de la ciudad, con alta incidencia de grietas de fatiga y deformaciones atribuibles a la combinación de suelos blandos, precipitaciones frecuentes y tráfico de carga informal.

La localidad de Puente Aranda, colindante al norte y nororiente del área de estudio, concentra actividad industrial y de servicios automotrices. Sus vías locales y colectoras registran deterioro asociado tanto al tráfico pesado como a las variaciones térmicas

características del altiplano bogotano. Los tramos que se extienden sobre la Avenida 1 de Mayo y la Transversal 53 actúan como corredores de articulación entre ambas localidades, con características de operación vial mixta.

A nivel de barrio, el sector de intervención corresponde principalmente al barrio San Eusebio y sus áreas limítrofes. Este barrio se encuentra delimitado por la Avenida 1 de Mayo al sur, la Carrera 52 al oriente y la Transversal 53 al occidente. Presenta una estructura urbana consolidada de manzanas ortogonales con alta densidad predial, lotes destinados a vivienda, comercio de escala barrial y algunos equipamientos de servicio urbano como el Parque Ferrocarril Tejar del León y la Clínica del Vestido. Las vías locales del sector son, en su mayoría, de calzada sencilla con anchos entre 6 y 9 metros, sin separador y con infraestructura de andenes y bordillos en estado variable.

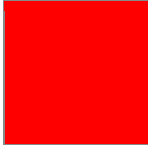
La selección de este sector como zona de captura se justifica por la heterogeneidad de la condición superficial de sus pavimentos, que permite obtener imágenes representativas de los tres niveles de deterioro (alto, medio y bajo) según la escala PCI, así como de las principales tipologías de grieta definidas en la norma ASTM D6433-20: grietas longitudinales, transversales, de bloque, de cocodrilo y grietas de borde. Esta diversidad es condición necesaria para garantizar la representatividad del conjunto de entrenamiento del modelo.).

Convención de colores — Clasificación PCI según ASTM D6433

En las figuras de localización de cada tramo, los segmentos evaluados se representan mediante una codificación cromática que refleja el nivel de deterioro superficial conforme al Índice de Condición del Pavimento (PCI), calculado según la norma ASTM D6433-20 adoptada por el IDU en la Guía GU-IC-15. La Tabla 4 resume la correspondencia entre color, nivel de deterioro y rango PCI aplicada en las figuras de este numeral:

Tabla 4.

Convención cromática de deterioro por rango PCI — ASTM D6433

Color	Nivel de deterioro	Clasificación PCI — ASTM D6433
	Rojo — Alto	PCI 0 a 40: deterioro alto. Comprende las categorías "Fallado" (0–10), "Muy malo" (11–25) y "Malo" (26–40). Requiere intervención estructural inmediata o rehabilitación mayor.
	Amarillo — Medio	PCI 41 a 70: deterioro moderado. Abarca "Regular" (41–55) y "Aceptable bajo" (56–70). Corresponde a mantenimiento preventivo o correctivo de mediana urgencia.
	Verde — Bajo	PCI 71 a 100: pavimento en buen estado. Comprende "Bueno" (71–85) y "Muy bueno / Excelente" (86–100). Solo requiere mantenimiento rutinario.

Nota. Elaboración propia con base en ASTM D6433-20 e IDU (2021).

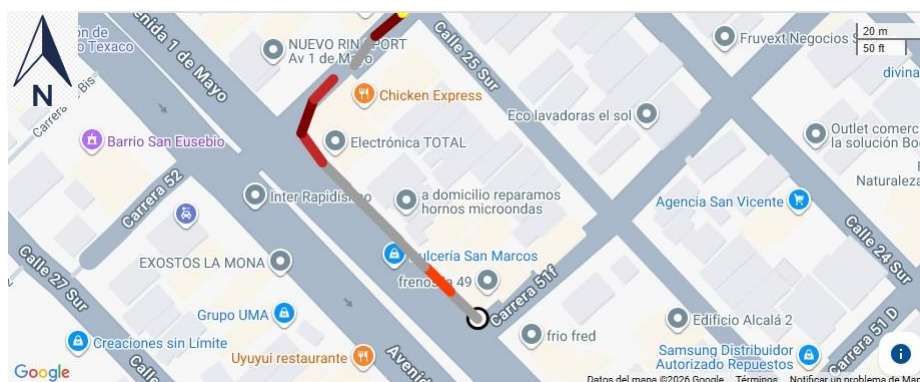
Localización específica de los tramos evaluados

A continuación, se describen los cuatro tramos sobre los cuales se realizó la captura de imágenes para predicción del modelo YOLOv26, con indicación de su localización, longitud aproximada y caracterización del estado superficial observado según la convención cromática definida en la Tabla 6:

Tramo 1 — Avenida 1 de Mayo (entre Carrera 51f y Carrera 52). Corredor arterial orientado en sentido suroccidente-nororiente, con calzada bidireccional y longitud evaluada aproximada de 180 metros. La Figura 15 muestra la localización del tramo. En la zona de intersección con la Carrera 52 se identifican segmentos con deterioro alto (rojo), asociados a grietas de cocodrilo y deformaciones superficiales producto del tráfico acumulado y la presencia de suelos blandos. Los sectores intermedios presentan condición de deterioro medio (amarillo), con grietas longitudinales de severidad moderada.

Figura 15.

Localización Tramo 1 — Avenida 1 de Mayo.

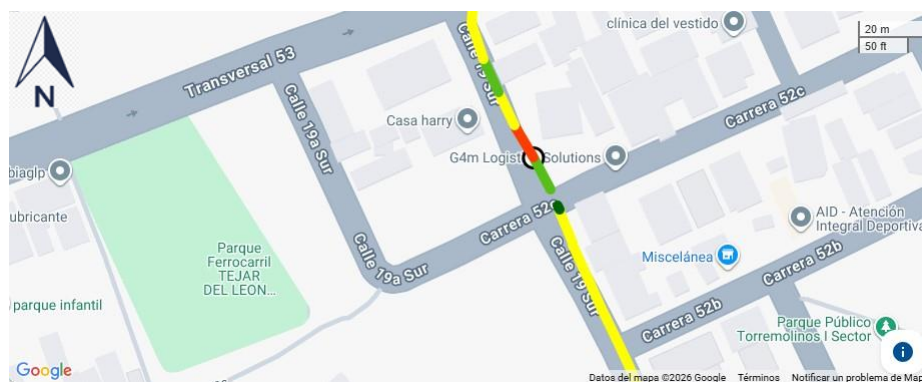


Nota. Fuente: elaboración propia con base en Google Maps © 2026.

Tramo 2 — Calle 19 Sur (entre Transversal 53 y Carrera 52). Vía local orientada en sentido occidente-oriente, con longitud evaluada de aproximadamente 220 metros. Según la Figura 16, el corredor presenta sectores con deterioro bajo (verde) en el extremo sur, en las inmediaciones del Parque Ferrocarril Tejar del León, y sectores con deterioro alto (rojo) concentrados en la zona de cruce frente al conjunto logístico G4m Logistic Solutions, donde se evidencian grietas longitudinales, transversales y en malla de severidad alta.

Figura 16.

Localización Tramo 2 — Calle 19 Sur

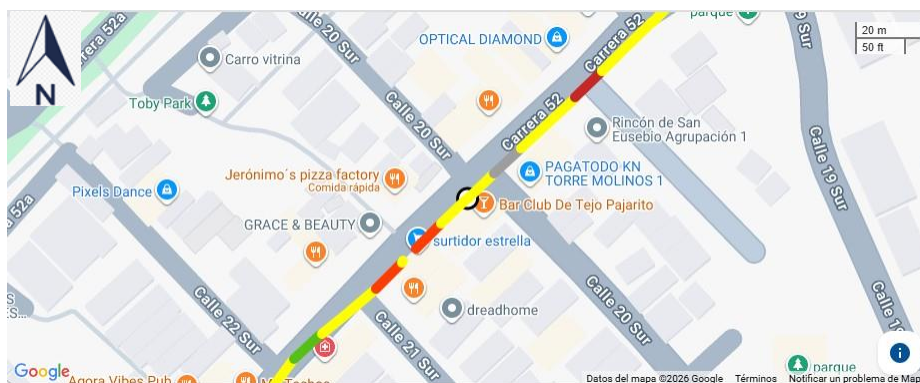


Nota. Fuente: elaboración propia con base en Google Maps © 2026.

Tramo 3 — Carrera 52 (entre Calle 20 Sur y Calle 22 Sur). Vía colectora con circulación vehicular de mediana intensidad, orientada en sentido norte-sur, con longitud evaluada de aproximadamente 300 metros. De acuerdo con la Figura 17, el tramo exhibe la mayor heterogeneidad de condición del conjunto: en el extremo norte, próximo a la Calle 20 Sur, se registran segmentos con deterioro alto (rojo) con grietas de bloque y fatiga; en la zona media, correspondiente a los sectores del Bar Club de Tejo Pajarito y Surtidor Estrella, se presentan sectores con deterioro medio (amarillo) con fisuras longitudinales de severidad moderada; y en el extremo sur se identifican sectores con deterioro bajo (verde) con superficie en condición aceptable.

Figura 17.

Localización Tramo 3 — Carrera 52



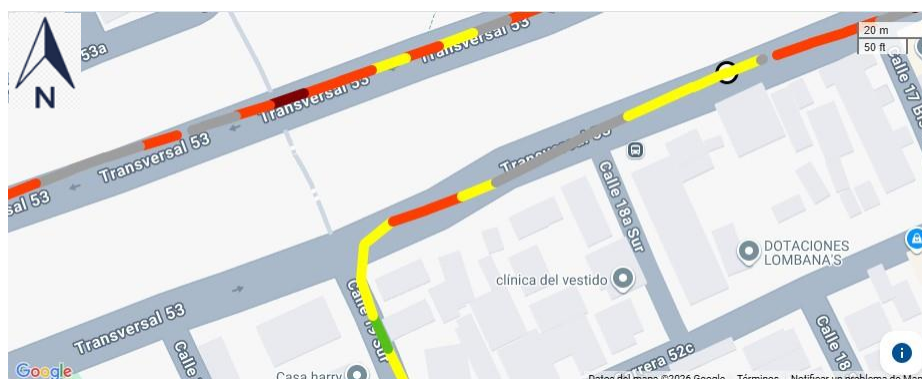
Nota. Fuente: elaboración propia con base en Google Maps © 2026.

Tramo 4 — Transversal 53 (entre Calle 18a Sur y Calle 19a Sur). Corredor local con orientación nororiente-suroriente, con longitud evaluada aproximada de 350 metros, que atraviesa dos ramales de la Transversal 53 y actúa como vía de conexión entre los tramos anteriores. Conforme a la Figura 18, el tramo presenta una distribución alternada de los tres niveles de condición: deterioro alto (rojo) en los extremos oriente y occidente del ramal principal, asociados a zonas de cruce vehicular de alta frecuencia; deterioro medio (amarillo) en el tramo recto central; y un segmento con deterioro bajo (verde) identificado en la curva de

conexión hacia la Calle 19 Sur, en las inmediaciones de la Clínica del Vestido. La coexistencia de los tres niveles en un solo corredor lo convierte en el tramo con mayor riqueza tipológica del conjunto evaluado.

Figura 18.

Localización Tramo 4 — Transversal 53.



Nota. Fuente: elaboración propia con base en Google Maps © 2026.

En conjunto, los cuatro tramos cubren una longitud total aproximada de 1.050 metros de vía local y colectora, con representación de los tres niveles de condición superficial según la escala PCI y de las principales tipologías de grieta definidas en la norma ASTM D6433-20. Esta diversidad garantiza que el conjunto de imágenes de entrenamiento y predicción incluya ejemplos suficientes para el adecuado entrenamiento y evaluación del modelo YOLOv26 implementado en la presente investigación.

Parámetros Modelo Deep Learning

El modelo utilizado de Yolo 26 tiene varias versiones de acuerdo con el tamaño de la red neuronal, por lo que para este caso de estudio se escogió la versión nano o la más pequeña.

A continuación, se presenta una tabla de resumen de sus parámetros y capas, considerando que varias de ellas corresponden a desarrollos personalizados de Ultralytics con

el propósito de mejorar el desempeño en los problemas de visión de computadora, correspondiente a 2.504.190 parámetros y 260 capas:

Tabla 5.

Resumen de parámetros Red YOLO V26 nano

Módulo	Argumentos	Parámetros
ultralytics.nn.modules.conv.Conv	[3, 16, 3, 2]	464
ultralytics.nn.modules.conv.Conv	[16, 32, 3, 2]	4672
ultralytics.nn.modules.block.C3k2	[32, 64, 1, False, 0.25]	6640
ultralytics.nn.modules.conv.Conv	[64, 64, 3, 2]	36992
ultralytics.nn.modules.block.C3k2	[64, 128, 1, False, 0.25]	26080
ultralytics.nn.modules.conv.Conv	[128, 128, 3, 2]	147712
ultralytics.nn.modules.block.C3k2	[128, 128, 1, True]	87040
ultralytics.nn.modules.conv.Conv	[128, 256, 3, 2]	295424
ultralytics.nn.modules.block.C3k2	[256, 256, 1, True]	346112
ultralytics.nn.modules.block.SPPF	[256, 256, 5, 3, True]	164608
ultralytics.nn.modules.block.C2PSA	[256, 256, 1]	249728
torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']	0
ultralytics.nn.modules.conv.Concat	[1]	0
ultralytics.nn.modules.block.C3k2	[384, 128, 1, True]	119808
torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']	0
ultralytics.nn.modules.conv.Concat	[1]	0
ultralytics.nn.modules.block.C3k2	[256, 64, 1, True]	34304
ultralytics.nn.modules.conv.Conv	[64, 64, 3, 2]	36992
ultralytics.nn.modules.conv.Concat	[1]	0
ultralytics.nn.modules.block.C3k2	[192, 128, 1, True]	95232
ultralytics.nn.modules.conv.Conv	[128, 128, 3, 2]	147712
ultralytics.nn.modules.conv.Concat	[1]	0
ultralytics.nn.modules.block.C3k2	[384, 256, 1, True, 0.5, True]	463104
ultralytics.nn.modules.head.Detect	[1, 1, True, [64, 128, 256]]	241566

Nota: Elaboración propia

A continuación, se describen las características de las capas, considerando la documentación del modelo Open Source disponible en (Ultralytics, 2026):

- **Conv:** En esta capa se encapsula una convolución 2D, Batch Normalization (Normalización de lotes) y una función de activación que predeterminadamente es SiLU (Unidad Lineal Sigmoide), siendo la ventaja de esta última tener una suavidad del salto entre encendido debido a la transformación sigmoidea de la entrada.
- **C3k2:** Corresponde a la capa que realiza el flujo de información entre los componentes de Backbone y Neck de las redes YOLO por medio del paradigma Cross Stage Partial (CSP), diseñado para balancear el flujo del gradiente y los costos computacionales. Adicional, permite diferentes tamaños de Kernel y estructuras de cuello de botella (Bottleneck).
- **SPPF:** Este bloque reúne propiedades de diferentes fuentes utilizando max-pooling con un tamaño de kernel fijo para reducir la latencia del cálculo.
- **C2PSA:** Variación del bloque de CSP introducido en YOLOv8 que utiliza el método Pointwise Spatial Attention (PSA) para mejorar el modelado del contexto global.
- **Concat:** Corresponde a capas de conexión entre los distintos bloques de YOLO
- **Head:** Capa de detección para obtener un valor asociado con regresión y clasificación basado en las Cajas de Frontera (Bounding Box).
- **Upsample:** Capa nativa de pytorch que realiza el proceso de aumento o upsampling a un tensor, sea de dimensionalidad (2D,3D) o de tamaño.

Las variables predominantes en el problema de investigación son de tipo técnico, dado que el estudio se centra en el análisis del desempeño de modelos de visión por computadora basados en Deep Learning para la detección de grietas en pavimentos flexibles, evaluando su precisión frente al método tradicional de inspección visual.

Si bien en el planteamiento del problema se identifican variables contextuales de carácter humano —como la subjetividad del evaluador— y variables operativas —como los tiempos de ejecución y los costos asociados al levantamiento de información—, estas no constituyen el

objeto principal de análisis, sino que representan condiciones del entorno en el que se aplica el método convencional y que motivan la búsqueda de alternativas tecnológicas.

Las variables técnicas que orientan el estudio incluyen:

- El tipo de modelo de visión por computadora utilizado que en este caso es Yolov26 (clasificación, detección de objetos, segmentación semántica).
- **Las métricas de evaluación empleadas:** precisión, F-measure, IoU y otros indicadores cuantitativos propios del campo de la inteligencia artificial.
- La capacidad de detección automática de grietas a nivel de píxel.
- El desempeño de los algoritmos de Deep Learning frente a condiciones variables de iluminación, textura superficial y heterogeneidad del material asfáltico.
- La compatibilidad metodológica entre la segmentación digital y la clasificación de deterioros por severidad definida en la norma ASTM D6433.

Como referencia de valores esperados, se consideran los resultados asociados con el entrenamiento de grietas del set de entrenamiento, porque debido a la naturaleza del problema los benchmarks de la herramienta disponibles en línea no son los adecuados, ya que el modelo fue entrenado en el dataset COCO (80 categorías), que tiene amplia variedad de objetos y que afecta las cifras finales:

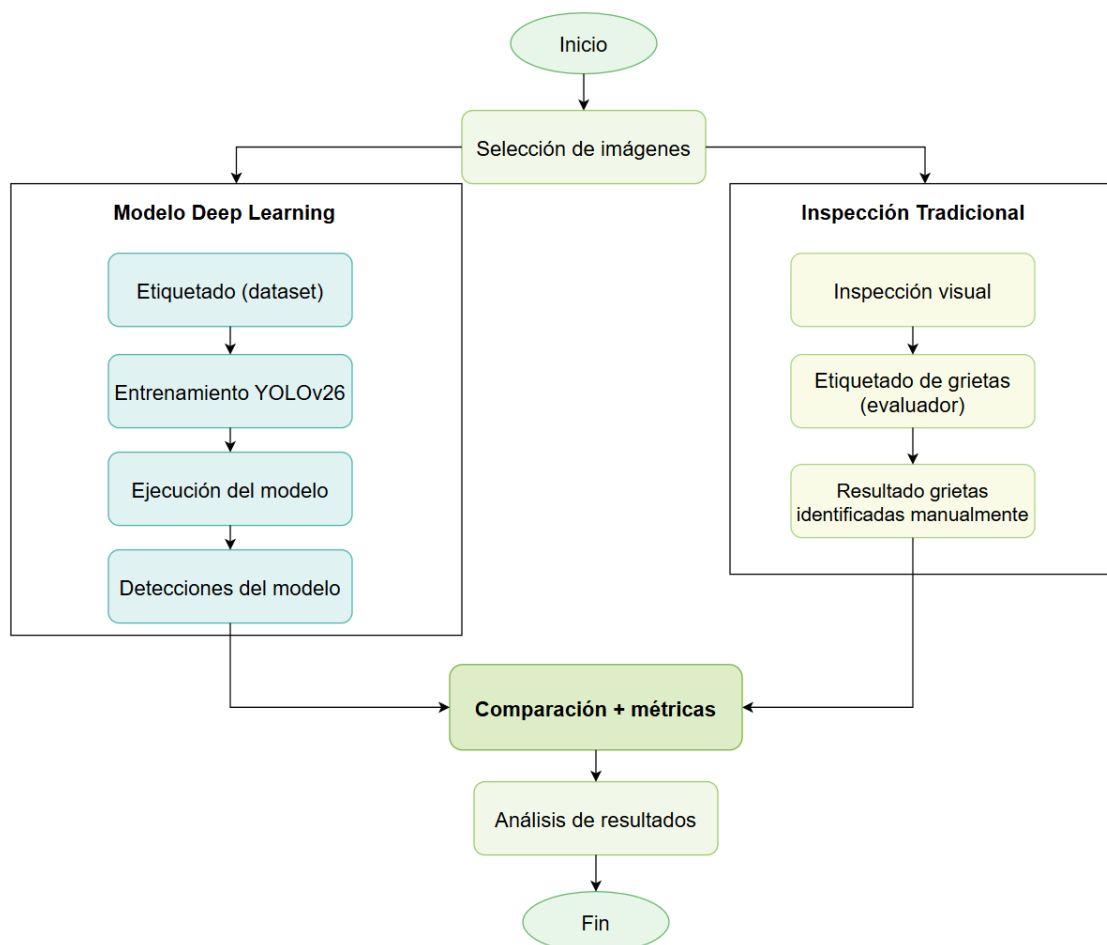
Tabla 6.

Valores esperados de desempeño

P	R	mAP50	mAP50-95
0.4 - 0.7	0.4 - 0.7	0.2 - 0.3	0.2 - 0.3

Nota: Elaboración propia

Finalmente, se presenta en la Figura 19 el resumen del flujo del proyecto:

Figura 19.*Diagrama de flujo del proyecto*

Nota: Elaboración propia

Resultados

Dataset de referencia

En primera instancia se presentan las imágenes del dataset de referencia, en las que se evidencia —a través de la Figura 20— la diversidad de casuísticas y entornos capturados. Por ejemplo, en la imagen de la parte inferior derecha se observa una sola grieta con orientación girada, correspondiente a un ejercicio de data augmentation dentro del conjunto de entrenamiento.

Por su parte, en la imagen inferior izquierda se aprecia la sombra de un dron durante la captura, elemento que puede influir en el desempeño del modelo. Estas imágenes resultan adecuadas para el proceso de detección de grietas en pavimentos flexibles, dado que recogen distintos ángulos, morfologías de grieta y tipos de superficie.

Figura 20.

Muestra de grietas en set de entrenamiento



Nota: Elaboración propia

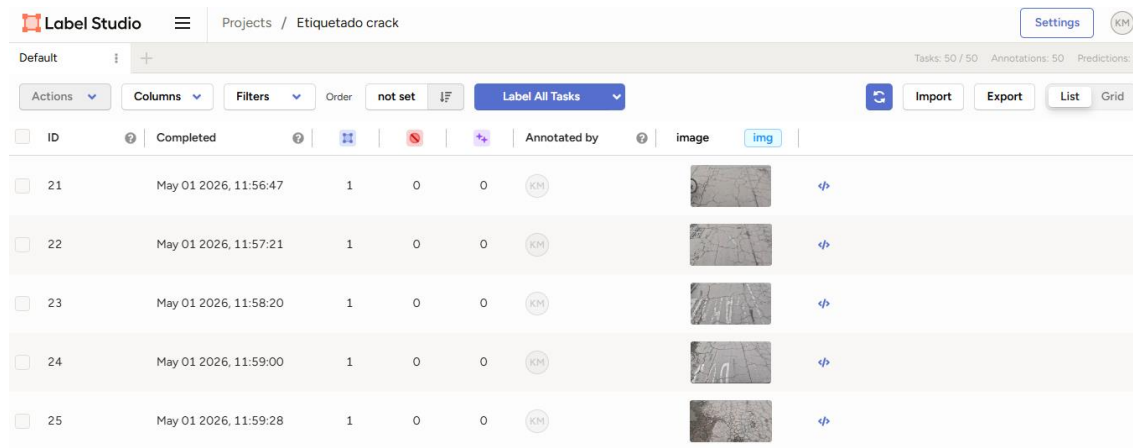
El total corresponde a 3.000 imágenes en distintos ángulos y rotaciones. Para el cálculo de las métricas de precisión, recall y mAP50-95 se utiliza el conjunto completo del dataset público, mientras que la validación se realiza exclusivamente con las imágenes de grietas reales capturadas en la ciudad de Bogotá.

Etiquetado

El etiquetado de las imágenes del tramo propuesto se realizó con la herramienta de software libre Label Studio, disponible como librería de Python y que genera un entorno web para el etiquetado de imágenes en distintas modalidades: clasificación, segmentación, detección de objetos (OBB), entre otras. La Figura 21 presenta la interfaz resultante aplicada a un conjunto de imágenes de grietas.

Figura 21.

Interfaz de etiquetado de imágenes



Nota: Elaboración propia

Desde la plataforma se cargan las imágenes y, una vez completado el etiquetado, se exportan en formato YOLO para su uso con la librería Ultralytics, seleccionando la versión más reciente: YOLOv26.

A lo largo del tramo escogido se identificaron distintos tipos de grietas. La Figura 22 muestra un primer caso en el que confluyen grietas interconectadas —con una fractura principal en el centro— junto con elementos adicionales como charcos de agua y el borde de un andén.

Figura 22.

Muestra de grietas #1



Nota: Elaboración propia

La Figura 23 ilustra un caso con menor densidad de grietas, concentradas en la parte derecha de la imagen y menos visibles que en el caso anterior. Se añade la presencia de señalización vial, elemento que puede confundirse con una grieta o reducir la capacidad del modelo para detectarlas. La Figura 24 presenta un comportamiento similar, con la incorporación de reductores de velocidad como elemento de la malla vial, acompañados de una grieta claramente visible en el lado derecho de la imagen.

Figura 23.

Muestra de grietas #2



Nota: Elaboración propia

Figura 24.

Muestra de grietas #3

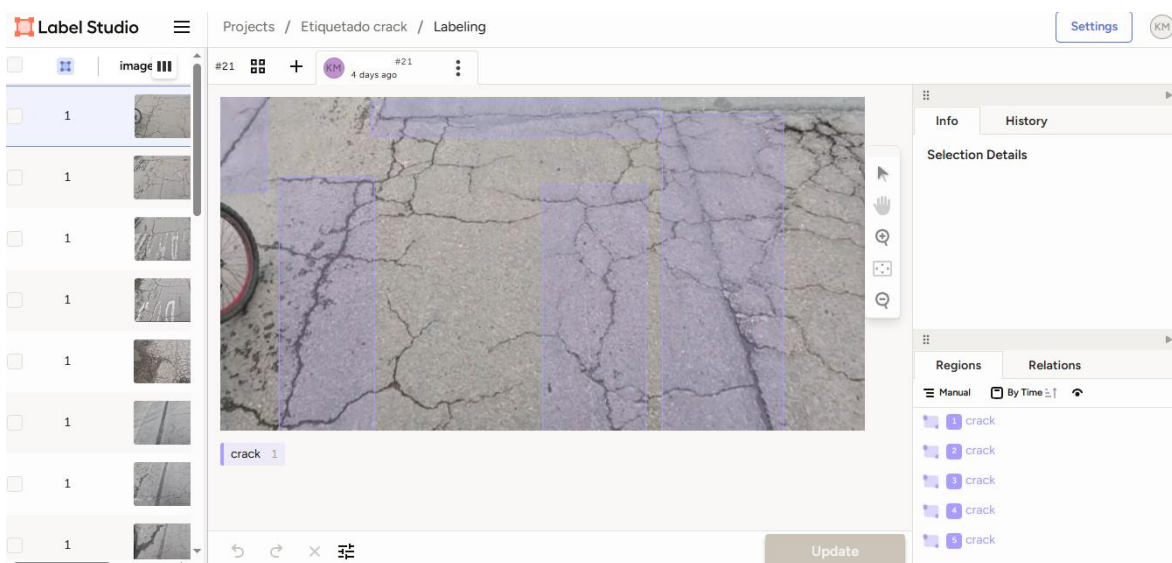


Nota: Elaboración propia

En la plataforma se etiquetó una única categoría denominada crack, nombre que coincide con el utilizado en el dataset de entrenamiento. La Figura 25 presenta un ejemplo real en el que, ante la presencia de varias grietas interconectadas, se optó por agruparlas siguiendo el mismo criterio del conjunto de entrenamiento, con el fin de evaluar cómo influye este tipo de daño en los valores finales de predicción.

Figura 25.

Etiquetado de grietas en pavimento en Label-studio



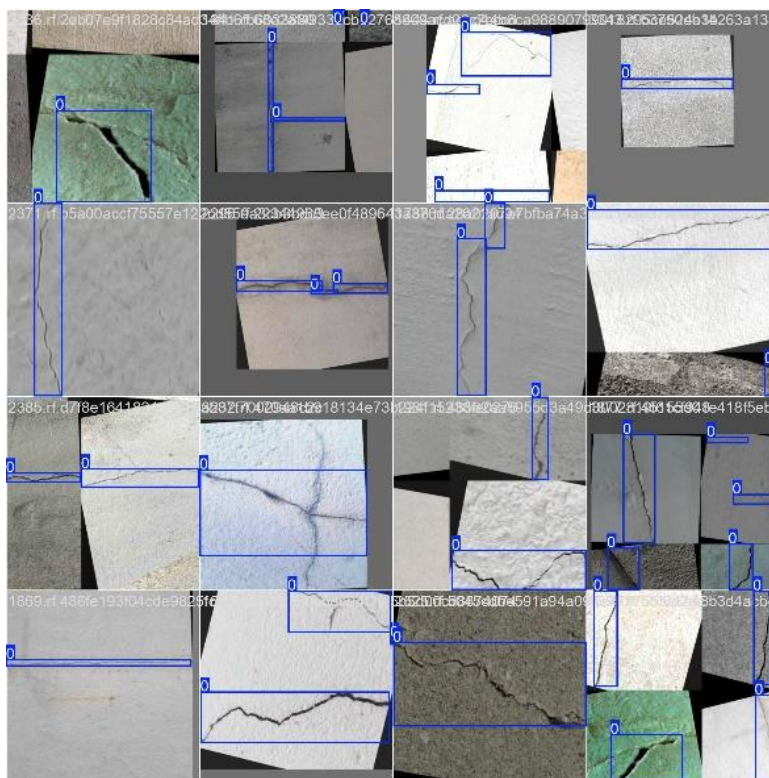
Nota: Elaboración propia

Entrenamiento modelo base

Con esto en cuenta, se lleva a cabo el entrenamiento del modelo sobre el conjunto de imágenes definido. La Figura 26 ilustra una muestra representativa de las imágenes junto con las cajas de contorno asignadas durante el etiquetado.

Figura 26.

Muestra de grietas de entrenamiento



Nota: Elaboración propia

El entrenamiento se ejecuta en un entorno de Google Colab, que cuenta con las principales librerías de aprendizaje de máquina y facilita la instalación de librerías adicionales como Ultralytics YOLO, con la consideración de que los archivos deben cargarse nuevamente en cada sesión. A continuación, se detallan las características del entorno de ejecución:

- Ultralytics 8.4.45
- Python 3.12.13

- torch 2.10.0+cu128
- GPU Tesla T4, 14.913 MiB

Se implementa el modelo nano de la familia YOLOv12 en la modalidad de detección de objetos, identificado como yolo12n y construido sobre PyTorch. Los parámetros de entrenamiento configurados fueron los siguientes:

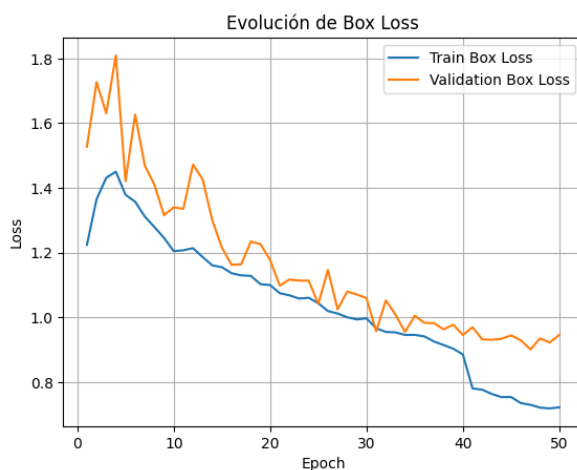
- Épocas: 50
- Tamaño de imagen: 512 px
- Caché de imágenes activado
- Reducción del aumento de datos (*data augmentation*), configurado en la función train
- AMP (*Automatic Mixed Precision*) para acelerar el entrenamiento

Tras aproximadamente una hora de entrenamiento, se obtuvieron los resultados que se presentan a continuación.

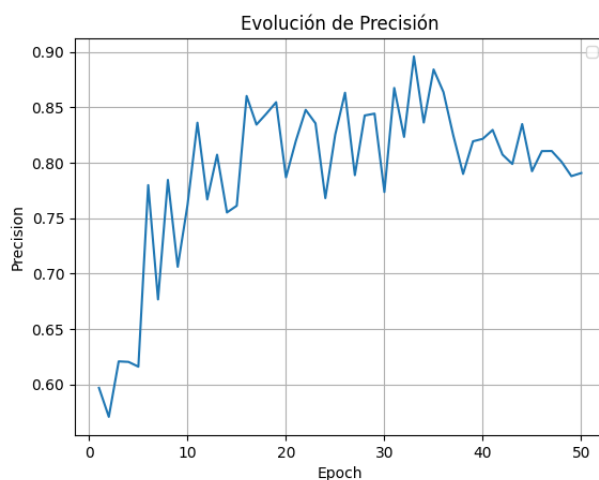
Gráficas de desempeño

La Figura 27 muestra la evolución de la pérdida en la función de costo asociada a las cajas de contorno. Salvo el pico inicial en las primeras cinco épocas, la curva desciende de forma sostenida, con un comportamiento similar al del conjunto de validación. A partir de la época 40 se observa una separación entre ambas curvas, lo que sugiere un inicio de sobreajuste (*overfitting*) al conjunto de entrenamiento. Adicionalmente, se aprecia una leve oscilación en la pérdida del conjunto de validación a lo largo del proceso.

En cuanto a la precisión, calculada sobre el conjunto de validación, la Figura 28 muestra una evolución con oscilaciones que alcanza un máximo del 90%, cerrando en un valor del 80% en la época 50. Consistente con lo observado en la curva de pérdida, el desempeño comienza a decrecer a partir de la época 40, lo que refuerza la hipótesis de sobreajuste al conjunto de entrenamiento.

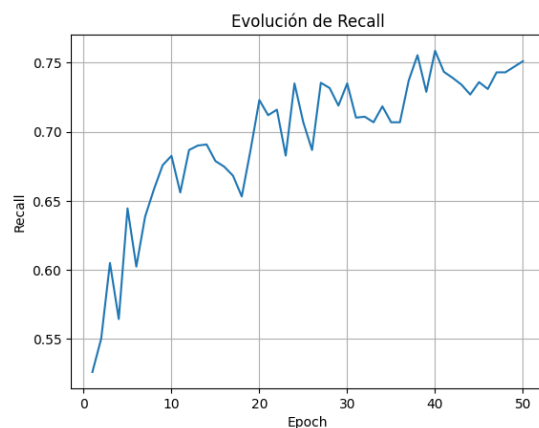
Figura 27.*Gráfica de pérdida*

Nota: Elaboración propia

Figura 28.*Evolución de precisión*

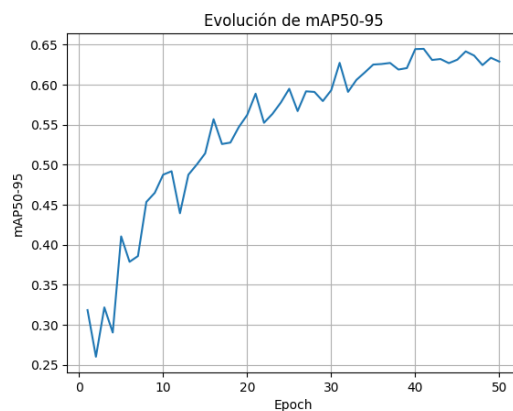
Nota: Elaboración propia

Por su parte, el recall presenta un incremento sostenido a lo largo de las épocas, con las oscilaciones propias del proceso, como se aprecia en la Figura 29, lo que indica una mejora progresiva en la capacidad del modelo para detectar grietas.

Figura 29.*Gráfica de evolución de recall*

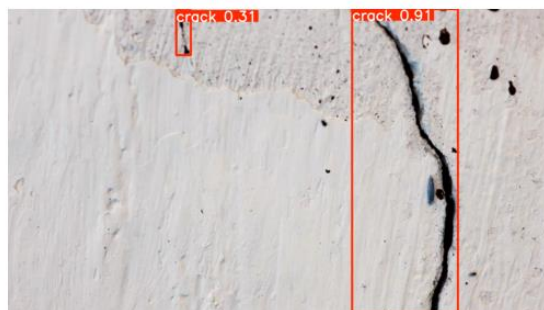
Nota: Elaboración propia

La Figura 30 presenta la métrica mAP50-95, que sigue la misma tendencia ascendente, confirmando que el modelo continúa aprendiendo a medida que avanzan las épocas.

Figura 30.*Métrica mAP50-95*

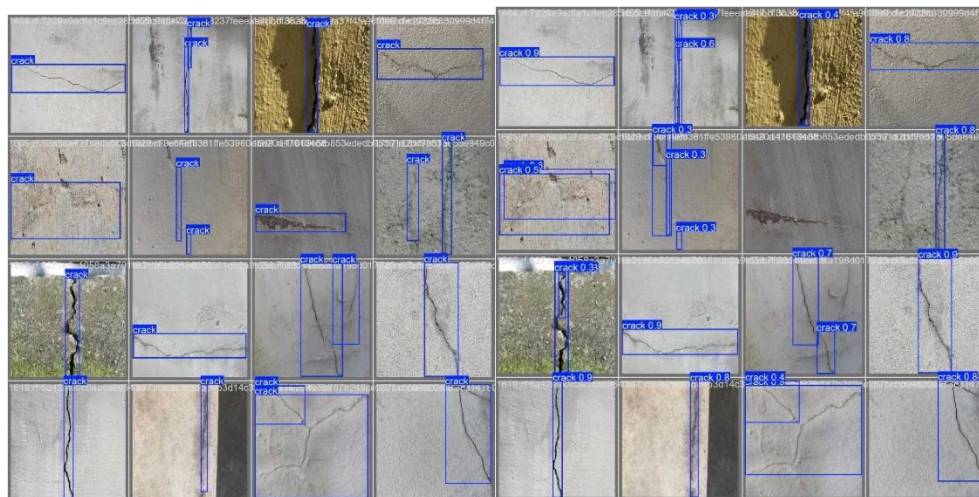
Nota: Elaboración propia

Finalmente, la Figura 31 ilustra un ejemplo de inferencia del modelo, en el que cada detección se representa mediante un cuadro de contorno acompañado de su puntuación de confianza, expresada en un rango de 0 a 1.

Figura 31.*Ejemplo de clasificación de grietas*

Nota: Elaboración propia

Un análisis más detallado se presenta en la Figura 32, que muestra una muestra de imágenes del conjunto de validación: a la izquierda la etiqueta original (ground truth) y a la derecha la predicción del modelo. Se observa que el modelo logra detectar correctamente la mayoría de las grietas, aunque presenta dificultades con aquellas de morfología menos definida.

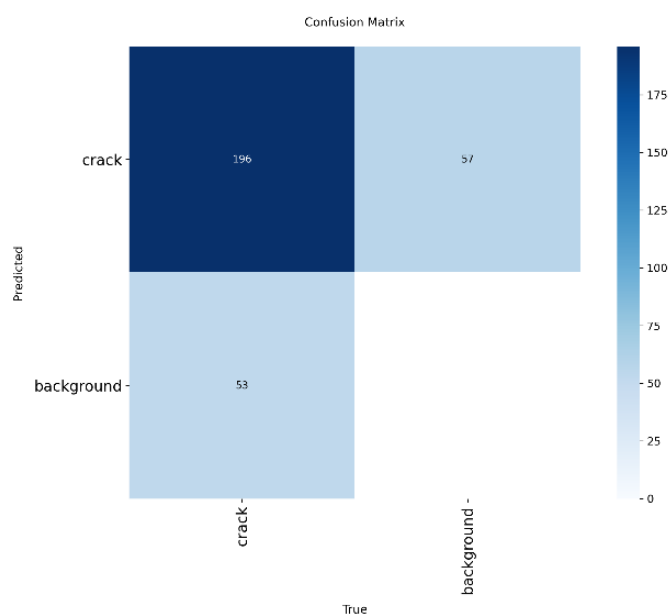
Figura 32.*Desempeño en conjunto de imágenes de entrenamiento*

Nota: Elaboración propia

Este resultado se complementa con la matriz de confusión de la Figura 33, en la que se aprecia que la mayor parte de las grietas es clasificada correctamente. Los errores se distribuyen entre falsos negativos —grietas no detectadas— y falsos positivos —superficies clasificadas como grieta en ausencia de daño—, representados en la matriz bajo la etiqueta background.

Figura 33.

Matriz de confusión entrenamiento



Nota: Elaboración propia

El modelo resultante fue guardado en formato PyTorch para su posterior evaluación sobre el conjunto de imágenes de grietas reales.

Validación con imágenes de grietas

Concluido el entrenamiento, se procedió a evaluar el modelo sobre las imágenes reales del tramo seleccionado. La Figura 34 presenta una muestra de las anotaciones de validación, en la que se aprecian los distintos tipos de grietas identificados a lo largo del recorrido.

Figura 34.*Muestra de marcación real*

Nota: Elaboración propia

A partir de estas imágenes se obtuvieron las predicciones del modelo, observándose que este tiende a detectar grietas de forma individual, sin agrupar las conexiones entre ellas, a diferencia del criterio utilizado durante el etiquetado. Este comportamiento se ilustra en la Figura 35.

Figura 35.*Predicciones del modelo Yolo en Grietas*

Nota: Elaboración propia

Los resultados arrojan una precisión de 0,442 y un recall de 0,404 como se muestra en la Figura 36. Si bien estos valores son inferiores a los obtenidos durante el entrenamiento, el modelo demuestra una velocidad de inferencia de pocos milisegundos por imagen, lo que lo posiciona como una referencia válida para aplicaciones en tiempo real.

Figura 36.

Predicciones del modelo Yolo en Grietas

```

Ultralytics 8.4.48 Python-3.12.13 torch-2.10.0+cpu CPU (Intel Xeon CPU @ 2.20GHz)
Downloading https://ultralytics.com/assets/Arial.ttf to '/root/.config/Ultralytics/Arial.ttf': 100% — 755.1KB 33.0MB/s 0.0s
val: Fast image access (ping: 0.0±0.0 ms, read: 780.7±435.1 MB/s, size: 54.8 KB)
val: Scanning /content/grietas_bogota/labels/val... 83 images, 0 backgrounds, 0 corrupt: 100% — 83/83 546.8it/s 0.2s
val: New cache created: /content/grietas_bogota/labels/val.cache
Class      Images  Instances  Box(P  R      mAP50  mAP50-95): 100% — 6/6 1.7s/it 10.2s
  all         83       274       0.442  0.404  0.364  0.172
Speed: 3.2ms preprocess, 109.1ms inference, 0.0ms loss, 0.2ms postprocess per image

```

Nota: Elaboración propia

Comparación con investigaciones anteriores

En primera instancia, como se presenta en la Figura 37 las grietas utilizadas en el dataset de entrenamiento corresponden a imperfecciones lineales, mientras que en las imágenes obtenidas de las calles en Bogotá tienen diferentes tipos.

Figura 37.

Muestras de grietas en entrenamiento - comparativa



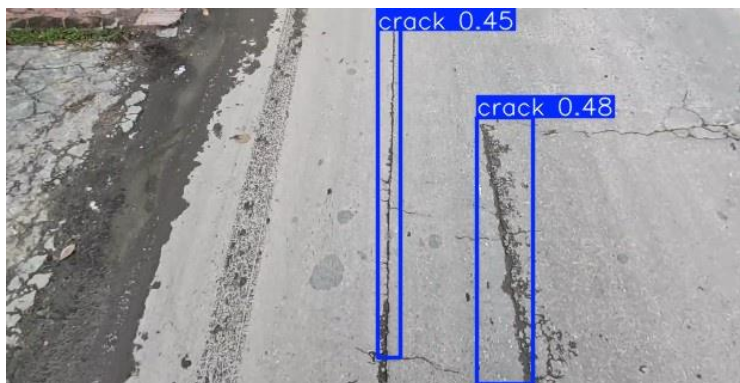
Nota: Elaboración propia

Por ejemplo, en la Figura 38 se evidencia que el modelo YOLO entrenado logró detectar las grietas verticales, que son similares a las del dataset de entrenamiento. De la misma

manera, en la Figura 39 se identifica que también detecta las grietas horizontales mientras que las demás imperfecciones de la vía, como lo son otro tipo de grietas interconectadas no son detectadas.

Figura 38.

Muestra de predicción de grietas verticales



Nota: Elaboración propia

Figura 39.

Muestra de predicción de grietas horizontales y verticales



Nota: Elaboración propia

El efecto de estas condiciones no se ven reflejados en las métricas obtenidas de entrenamiento con respecto a las investigaciones de (Li et al., 2025) con el modelo YOLOv5 adaptado a análisis de fisuras en piedras y el modelo YOLO - LY para detección de grietas en puentes de (Yin et al., 2026) mostrado en la Tabla 7, donde las métricas de Precisión, Recall y

mAP@50 son similares para estos dos casos más el modelo implementado. Puntualmente, la precisión entre YOLOv5 y YOLOv26 es similar mientras que destaca que el Recall de este último es el menor, debido a que no se detectaron correctamente todas las grietas del set escogido para esta investigación, con un número relativamente alto de Falsos Negativos, lo que impactó la cifra:

Tabla 7.

Comparación de métricas para entrenamiento

Modelo	Precisión	Recall	mAP@50
YOLO – LY	0.903	0.777	0.833
YOLOv5	0.85	0.86	0.87
YOLOv26 – Dataset crack	0.849	0.721	0.81

Nota: Elaboración propia

Con respecto a las métricas con los datos reales o de validación, en las dos investigaciones de referencia se entrena el modelo con datasets creados de la misma información que se valida. Por ejemplo, de etiquetado de grietas en rocas o de grietas en puentes, mientras que en esta investigación el enfoque fue utilizar una técnica de transfer learning, que consistió en entrenar el modelo inicial con una base de más de 3 mil imágenes con etiquetas claras y transferirlas a un escenario similar, como lo es la detección de grietas en pavimento flexible. Estos resultados se presentan en la Tabla 8.

Tabla 8.

Comparación de métricas para validación

Modelo	Precisión	Recall	mAP@50
YOLO – LY*	0.903	0.777	0.833
YOLOv5*	0.85	0.86	0.87
YOLOv26 – Dataset crack	0.442	0.404	0.364

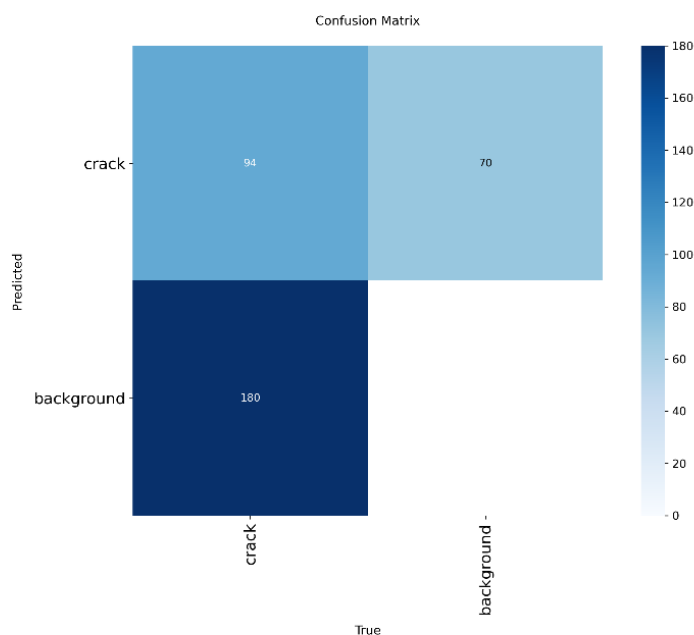
Nota: Elaboración propia

Para comparar, se colocó el mismo valor de entrenamiento en las métricas de validación de los modelos compartidos en la literatura, ya que en ambos no se presenta una comparación

tradicional entre desempeño de entrenamiento y validación como se realizó en esta investigación. Esto hace que los valores obtenidos en la literatura sean más altos que los de esta investigación, siendo en este caso la precisión de **0.442** o la mitad de los otros estudios, complementado con la matriz de confusión que se muestra en la Figura 40.

Figura 40.

Matriz de confusión predicción de grietas



Nota: Elaboración propia

Por último, la matriz de confusión permite visualizar el desempeño del modelo de detección, mostrando la cantidad de grietas reales detectadas correctamente, grietas reales no detectadas y zonas del pavimento clasificadas de manera incorrecta. Específicamente, se identifica que la mayor parte de grietas no son detectadas, correspondiente a la sección de Falsos Negativos o que fue predicho como background. A su vez, hay un número reducido en donde se predicen otros entornos del ambiente como grietas, en especial los bordes de los andenes.

Con ello, se concluye que este modelo es una referencia para aplicaciones futuras, ya que aún con un número reducido de imágenes se logró obtener una predicción cercana a la realidad con los retos que supone la identificación de grietas en los pavimentos de ciudad de Bogotá.

Conclusiones

Se identificó que los modelos de visión por computadora basados en Deep Learning más utilizados para detección de grietas corresponden principalmente a arquitecturas tipo YOLO, debido a su capacidad de detección en tiempo real y buen desempeño en tareas de visión por computadora. En este trabajo se utilizó YOLOv26 por sus mejoras frente a versiones anteriores.

Asimismo, se identificó que las métricas más utilizadas en estudios relacionados con detección de grietas corresponden a precisión, recall, IoU y mAP. Estas métricas permiten evaluar de manera cuantitativa el desempeño de los modelos de detección y facilitan la comparación frente a métodos tradicionales de inspección visual. Su uso en la investigación permitió analizar no solamente la capacidad del modelo para detectar grietas, sino también el nivel de coincidencia entre las detecciones automáticas y las referencias manuales definidas durante el proceso de etiquetado.

Durante el proceso de etiquetado realizado con Label Studio se observó que la forma en que se etiquetan las grietas influye directamente en el desempeño del modelo. En particular, agrupar grietas cercanas dentro de una misma etiqueta generó mejores resultados que utilizar múltiples etiquetas separadas para una misma zona de daño, donde la precisión tendía a disminuir. Esto permitió evidenciar la importancia de mantener criterios consistentes durante el etiquetado.

A su vez, la familia de modelos YOLO es utilizada ampliamente por su facilidad de entrenamiento y alto desempeño basado en la capacidad de detectar correctamente distintos tipos de objetos mientras se dispone de una librería en Python que facilita el uso de estas arquitecturas y se tiene la flexibilidad de transformar el tipo de detección desde la identificación de daños pixel a pixel o de ubicación de grietas.

Puntualmente, para poder lograr una precisión de más del 80% se utilizó un modelo de pocos parámetros como lo es la versión nano de YOLOv26 y con tiempo de ejecución de menos de 1 hora, lo que permitió utilizar un hardware no tan especializado y que se puede adaptar a varias aplicaciones. A su vez, la inferencia para nuevas imágenes es rápida aun en un entorno sin GPU, lo que facilita su implementación en entornos con recursos limitados.

Finalmente, la comparación entre el modelo basado en Deep Learning y la inspección visual tradicional permitió concluir que los modelos de visión por computadora representan una alternativa viable para apoyar la identificación de grietas en pavimentos flexibles. Sin embargo, su desempeño depende en gran medida de factores como la calidad de las imágenes, el proceso de entrenamiento y la forma en que se realiza el etiquetado de las grietas.

Durante el desarrollo del trabajo se evidenció que mantener criterios consistentes de etiquetado influye directamente en métricas como la precisión, por lo que estandarizar este proceso resulta importante para obtener resultados más confiables. De igual manera, contar con un dataset amplio y bien etiquetado permite mejorar la capacidad del modelo para reconocer distintos tipos de grietas y condiciones del pavimento, favoreciendo un desempeño más estable frente a la inspección visual tradicional.

Referencias

- ASTM. (2018). *D6433 - Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys*. <https://doi.org/10.1520/D6433-18>
- El-Din Hemdan, E., & Al-Atroush, M. E. (2025). A Review Study of Intelligent Road Crack Detection: Algorithms and Systems. *International Journal of Pavement Research and Technology* 2025, 1–31. <https://doi.org/10.1007/s42947-025-00556-x>
- Fakhri, M., Pourjafar, S. V., & Daneshvari, M. H. (2025). Texture-based image analysis and explainable machine learning for polished asphalt identification in pavement condition monitoring. *Scientific Reports* 2025 15:1, 15(1), 43167-. <https://doi.org/10.1038/s41598-025-27203-6>
- Hoang, N. D., & Tran, V. D. (2022). Computer vision based asphalt pavement segregation detection using image texture analysis integrated with extreme gradient boosting machine and deep convolutional neural networks. *Measurement*, 196, 111207. <https://doi.org/10.1016/j.measurement.2022.111207>
- Instituto de Desarrollo Urbano. (2021). *Metodología de diagnóstico y determinación del estado de los pavimentos y otros elementos del espacio público de Bogotá D.C.*
- Instituto de Desarrollo Urbano. (2025). *Diagnóstico, técnicas de intervención y diseño de alternativas para la conservación programada de los pavimentos de la malla vial, el espacio público y la red de ciclorrutas de Bogotá, D.C.*
- Jocher, G., & Qiu, J. (2026). *Ultralytics YOLO26*. <https://github.com/ultralytics/ultralytics>
- Li, W., Gao, Z., Feng, G., Hao, R., Zhou, Y., Chen, Y., Liu, S., Zhang, H., & Wang, T. (2025). Damage characteristics and YOLO automated crack detection of fissured rock masses

- under true-triaxial mining unloading conditions. *Engineering Fracture Mechanics*, 314, 110790. <https://doi.org/10.1016/J.ENGFRACMECH.2024.110790>
- Morshed, S. Y., & Bheemasetti, T. v. (2025). Classification and Severity Assessment of Pavement Cracks with Low-Cost Data Collection Systems and Machine Learning Models. *International Journal of Pavement Research and Technology* 2025, 1–14. <https://doi.org/10.1007/s42947-025-00680-8>
- Nhat-Duc, H., & Van-Duc, T. (2023). Computer Vision-Based Severity Classification of Asphalt Pavement Raveling Using Advanced Gradient Boosting Machines and Lightweight Texture Descriptors. *Iranian Journal of Science and Technology, Transactions of Civil Engineering* 2023 47:6, 47(6), 4059–4073. <https://doi.org/10.1007/s40996-023-01138-2>
- Ramírez-Villanueva, F. G., Vázquez Noguera, J. L., Legal-Ayala, H., Mello-Román, J. C., & Pérez-Estigarribia, P. E. (2025). PY-CrackDB: A pavement crack dataset from paraguayan roads for context-aware computer vision models. *Data in Brief*, 63, 112060. <https://doi.org/10.1016/j.dib.2025.112060>
- Tafida, A., Alaloul, W. S., Zawawi, N. A. B. W., Musarat, M. A., & Sani, A. A. (2025). Advancing smart transportation: A review of computer vision and photogrammetry in learning-based dimensional road pavement defect detection. *Computer Science Review*, 56, 100729. <https://doi.org/10.1016/j.cosrev.2025.100729>
- Tian, Y., Xu, W., Yang, B., Yang, X., Guo, H., Wang, G., & Yu, H. (2026). Development and evolution of YOLO in object detection: A survey. *Neurocomputing*, 669, 132436. <https://doi.org/10.1016/j.neucom.2025.132436>
- Ultralytics. (2026). *YOLO Model Architecture and Task Variants*. <https://deepwiki.com/ultralytics/ultralytics/3-model-operations-and-lifecycle>

- Yin, X., Wang, C., Chen, W., Zeng, Z., Quan, Y., & Huang, Z. (2026). A UAV-deployable lightweight framework for real-time bridge crack detection via YOLO-LY algorithm. *Engineering Structures*, 358, 122629. <https://doi.org/10.1016/J.ENGSTRUCT.2026.122629>
- Zeida, W., Alnaqbi, A. J., Ghazi, ·, Al-Khateeb, G., & Abuzwidah, · Muamer. (2024). Machine learning modeling of transverse cracking in flexible pavement. *Discover Civil Engineering* 2024 1:1, 1(1), 114-. <https://doi.org/10.1007/s44290-024-00128-1>
- Zeng, J., Gül, M., & Mei, Q. (2022). A computer vision-based method to identify the international roughness index of highway pavements. *Journal of Infrastructure Intelligence and Resilience*, 1(1), 100004. <https://doi.org/10.1016/j.iintel.2022.100004>
- Zhang, A. A., Ding, Y., Xu, X., Dong, Z., He, A., Zhan, Y., Qian, Y., & Wang, K. C. P. (2025). Pixel-Level Intelligent Recognition of Asphalt Pavement Cracks with an Improved ShuttleNet. *International Journal of Pavement Research and Technology* 2025, 1–15. <https://doi.org/10.1007/s42947-025-00573-w>
- Zhang, Y., Li, G., & Shen, R. (2026). Computer vision for infrastructure defect detection: Methods and trends. *Automation in Construction*, 183, 106795. <https://doi.org/10.1016/j.autcon.2026.106795>
- Zhang, Y., & Liu, C. (2025). Vision-enhanced multi-modal learning framework for non-destructive pavement damage detection. *Automation in Construction*, 177, 106389. <https://doi.org/10.1016/j.autcon.2025.106389>