

Sistema de Análisis de Imágenes para la Identificación y Evaluación de Agrietamientos y Baches en pavimentos flexibles

Autor: Juan Diego González Jiménez

Tutor: Ing. Manuel Santiago Ocampo Terreros
(Ingeniero Civil - Magíster en Ingeniería Civil - Doctor en Ingeniería)

Co-Tutor: Ing. Alejandro Castro
(Ingeniero Eléctrico - Magíster en Ingeniería de Sistemas y Computación)

Pontificia Universidad Javeriana
Facultad de Ciencias
Pregrado en Ciencia de Datos
Bogotá D.C
2026

Contenido

Agradecimientos **3**
Resumen **4**
1. Introducción **5**
2. Objetivo General **7**
2.1. Objetivos Específicos 7
3. Marco teórico **8**
3.1. Pavimentos y clasificación 8
3.2. Mezcla Asfáltica 12
3.3. Pavimentos en Bogotá 14
3.4. Fallas en el Pavimento 15
3.5. Técnicas de reparación 17
3.6. Trabajos Asociados 19
4. Metodología Propuesta **32**
4.1. Marco de Gestión del Proyecto 32
4.2. Diseño del Sistema Multi-Experto y origen de datos 38
5. Resultados Preliminares **39**
5.1. Configuración Experimental y Adquisición de Datos 40
5.2. Resultados del Preprocesamiento y Etiquetado 42
5.2.1. Pipeline de Preprocesamiento de Imágenes *42*
5.2.1. Limpieza y Definición de Etiquetas *43*
5.3. Primeros Resultados preliminares 45
6. Resultados Finales **52**
6.1. Evaluación Cuantitativa y Comparativa 54
6.2. Resultados Finales 60
6.3 Formulación Matemática del Sistema Multi-Experto 74
6.3.1 Definición del espacio para entrada y salida del modelo *74*
6.3.2 Optimización y Funciones de Pérdida *75*
6.3.3. Inferencia con aumento de datos en tiempo de prueba (TTA) *76*
6.3.4. Umbralización Dinámica y Fusión Multi-Experto *76*
7. Conclusiones **77**
8. Bibliografías **82**
Anexos: Repositorio de Datos y Modelos **89**

Agradecimientos

Quiero expresar mi más sincero agradecimiento a mi familia, por su apoyo incondicional a lo largo de este camino. Gracias por el esfuerzo constante, por el respaldo económico y, sobre todo, por el apoyo emocional en los momentos más difíciles. Su confianza en mí y su disposición para ayudarme cuando lo necesité fueron fundamentales para alcanzar este logro.

A mi novia, por estar siempre a mi lado, por ser un soporte en los momentos en los que quise rendirme y por impulsarme a seguir adelante. Gracias también por tu apoyo en la redacción y estructura de este documento, y por creer en mí incluso cuando yo dudaba.

A mis tutores, por su guía, paciencia y acompañamiento durante todo el proceso. Sus ideas, observaciones y correcciones fueron clave para fortalecer este trabajo, así como su disposición para enseñarme y orientarme en cada etapa.

A todos mis profesores y a las personas con las que compartí durante mi paso por la universidad, por las enseñanzas técnicas y personales que contribuyeron a mi formación integral.

Finalmente, también me reconozco a mí mismo por el esfuerzo, la constancia y la disciplina invertidos en este proceso. Este trabajo de grado no solo representa un requisito académico, sino un reto personal que me permitió aplicar lo aprendido, enfrentar desafíos y crecer tanto profesional como personalmente.

Resumen

Este trabajo de grado presenta el desarrollo e implementación de un sistema automatizado de análisis de imágenes para la identificación y evaluación de fallas en pavimentos flexibles urbanos, con énfasis en el contexto vial de Bogotá. El sistema propuesto se fundamenta en la arquitectura de red neuronal convolucional U-Net y adopta un enfoque Multi-Experto, en el que dos modelos especializados e independientes abordan de forma diferenciada la segmentación semántica de fisuras y baches, dada la marcada diferencia morfológica entre ambos tipos de daño.

Para el especialista en fisuras se implementó una Attention U-Net entrenada mediante Transfer Learning sobre un dataset externo de 7.732 imágenes proveniente de HuggingFace, seguido de fine-tuning con el dataset local de pavimento. Para el especialista en baches se empleó una U-Net con encoder VGG16 preentrenado en ImageNet, complementada con técnicas de Data Augmentation que expandieron el conjunto de entrenamiento de 568 a más de 8.500 iteraciones aumentadas. Ambos modelos fueron optimizados con funciones de pérdida diseñadas para contrarrestar el desbalance extremo de clases, combinando Focal Loss y Tversky Loss para fisuras, y BCE con Dice Loss para baches.

El sistema fue evaluado mediante métricas estándar de segmentación semántica y validado con una estrategia de validación cruzada K-Fold de cinco particiones. El especialista en baches alcanzó un IoU medio de 0.403 ± 0.036 y un F1 de 0.611 ± 0.035, mientras que el especialista en fisuras obtuvo un IoU de 0.186 ± 0.007, resultado coherente con los rangos reportados en la literatura para segmentación de estructuras lineales delgadas bajo alto desbalance de clases. La baja desviación estándar en ambos modelos confirma la ausencia de sobreajuste y la generalización del aprendizaje. Adicionalmente, el sistema integra Test-Time Augmentation, umbralización dinámica óptima y una clasificación de severidad por porcentaje de área afectada, generando un diagnóstico visual interpretable sobre imágenes capturadas en condiciones reales.

Palabras clave: Pavimentos flexibles; Segmentación semántica; U-Net; Sistema Multi-Experto; Detección de baches; Detección de fisuras; Transfer Learning; Validación cruzada; Visión por computadora; Malla vial.

1. Introducción

La congestión vehicular es una problemática mundial que afecta a los países en desarrollo, especialmente a las ciudades con una población en continuo crecimiento. Esto hace que las ciudades requieren vías óptimas y amplias para albergar la gran cantidad de vehículos de sus habitantes. Por ello, es fundamental que las ciudades mantengan una alta calidad en sus vías, sin embargo, esto no ocurre en la mayor parte del mundo.

Existen diversos factores que influyen en la fatiga del pavimento en las ciudades, siendo uno de los más relevantes la inestabilidad en las capas que conforman la estructura vial. En muchos casos, estos problemas se originan a partir de intervenciones previas en las que hubo una inadecuada estructuración de la vía o el uso de materiales de baja calidad. Además, no todos los agregados presentan el mismo comportamiento: algunos materiales triturados logran una mejor compactación y cohesión, lo que contribuye a una cimentación más resistente. Por ello, resulta fundamental seleccionar los agregados de acuerdo con las necesidades específicas de cada vía (Frapial, 2022).

Además, las condiciones climáticas propias de cada ciudad constituyen un factor determinante en la aparición de fallas en los pavimentos. Por ejemplo, durante el verano, la exposición a la radiación ultravioleta (UV), especialmente a los rayos UVB, acelera la degradación de los materiales asfálticos. Por su parte, en invierno, las bajas temperaturas pueden provocar el congelamiento de la superficie de la vía, lo que dificulta las labores de mantenimiento y puede aumentar las cargas sobre el pavimento debido a la acumulación de hielo o humedad.

En contraste, durante la primavera, las lluvias y los ciclos de congelación y descongelación posteriores al invierno favorecen la infiltración de agua hacia las capas inferiores del pavimento. Este proceso debilita su estructura interna y contribuye a la formación y propagación de deterioros como baches, grietas y el fenómeno conocido como “piel de cocodrilo”.

En consecuencia, el pavimento está sometido a variaciones climáticas a lo largo del año que generan cambios constantes en su entorno. Si no se emplean materiales adecuados desde su

construcción ni se realiza un mantenimiento oportuno, su deterioro puede acelerarse significativamente (NAC Supply, Inc., 2019).

Además de lo anterior, las fallas en el pavimento no solo afectan la infraestructura, sino que también tienen consecuencias sociales y económicas importantes. Por ejemplo, en Bogotá se estima que se necesitarán cerca de 11 billones de pesos para llevar a cabo la rehabilitación y el mantenimiento integral de la malla vial, una cifra bastante alta que refleja la magnitud del problema. Esta situación no es exclusiva de la capital colombiana, sino que se repite en muchas ciudades del mundo (Guzmán, 2025). El mantenimiento de las redes viales suele ser altamente costoso y, en muchos casos, poco eficiente, lo que ocasiona que las obras de reparación o conservación se extiendan durante largos periodos. En efecto, se produce un colapso en la movilidad urbana por el alto flujo vehicular, lo que deriva en pérdidas de tiempo por congestiones, accidentes, daños en los automotores e incluso en situaciones donde los ciudadanos ponen en riesgo sus vidas.

Es por esto que este trabajo de grado tiene como propósito identificar de manera temprana y precisa la presencia de grietas, fisuras, baches y fallas tipo “piel de cocodrilo” en la infraestructura vial, mediante técnicas de análisis de imágenes y machine learning. El objetivo es evitar que estas fallas evolucionen hacia problemas más severos, como hundimientos, que demanden mayores inversiones y afecten de forma más crítica la movilidad urbana. De igual manera, se busca disminuir la ocurrencia de accidentes, ya que intervenir las fallas en etapas tempranas requiere menos tiempo y recursos. Esto también contribuye a reducir los daños en los vehículos, cuyos costos de reparación suelen ser elevados. Además, una detección oportuna podría incluso salvar vidas, pues una vía en buen estado reduce la probabilidad de siniestros que afectan principalmente a motociclistas y ciclistas, quienes son los más vulnerables ante este tipo de deterioros.

2. Objetivo General

Desarrollar e implementar un sistema automatizado de análisis de imágenes basado en arquitecturas de segmentación semántica profunda que permita identificar, delimitar y evaluar fallas superficiales en pavimentos flexibles urbanos, específicamente fisuras y baches, mediante un enfoque Multi-Experto que genere un diagnóstico cuantificable del área afectada y el nivel de severidad del daño, con el fin de contribuir a la gestión preventiva del mantenimiento vial y reducir la dependencia de inspecciones manuales subjetivas.

2.1. Objetivos Específicos

- Diseñar e implementar un pipeline de adquisición y preprocesamiento de imágenes viales que garantice la estandarización, limpieza y correcta codificación de las máscaras de segmentación, habilitando el entrenamiento supervisado de los modelos especializados.

- Construir y entrenar dos arquitecturas U-Net independientes y especializadas, una orientada a la detección de fisuras mediante Attention U-Net con Transfer Learning, y otra orientada a la detección de baches mediante U-Net con encoder VGG16, incorporando funciones de pérdida adaptadas al desbalance de clases propio de cada tipo de falla.

- Evaluar el desempeño de cada modelo especialista mediante métricas cuantitativas de segmentación semántica, incluyendo IoU, Coeficiente Dice, Precisión, Recall y F1, aplicando validación cruzada K-Fold de cinco particiones para garantizar la robustez y generalización de los resultados.

- Implementar y validar cualitativamente el sistema Multi-Experto completo sobre imágenes capturadas en condiciones reales del entorno vial bogotano, verificando su

capacidad para detectar simultáneamente ambos tipos de falla, cuantificar el porcentaje de área afectada y clasificar la severidad del daño en cuatro niveles operativos.

3. Marco teórico

3.1. Pavimentos y clasificación

Un pavimento es una estructura construida sobre la superficie natural del terreno con el fin de soportar y distribuir las cargas producidas por el tránsito vehicular, garantizando comodidad, seguridad y durabilidad en la vía (Yoder & Witczak, 1975). Su función principal es transmitir de manera gradual los esfuerzos generados por el tráfico hacia el suelo de fundación. Sin embargo, existen distintos tipos de pavimentos que se implementan en las ciudades según factores como el volumen de tránsito, el tipo de vehículos que circulan, las condiciones climáticas, la disponibilidad de materiales y los costos de mantenimiento. Es por esto, que los pavimentos se clasifican en flexibles, semirrígidos, rígidos y articulados.

El pavimento flexible está conformado por varias capas granulares elaboradas con materiales pétreos, como roca triturada y grava de diferentes tamaños, dispuestas sobre una mezcla asfáltica. Estas capas trabajan en conjunto para distribuir las cargas de manera progresiva, reduciendo los esfuerzos gradualmente hasta alcanzar la subrasante. Este pavimento se usa en vías urbanas, carreteras de mediano y alto tráfico, por lo que tiene una alta capacidad de deformación, aunque es fácil su rehabilitación, es muy sensible a variaciones de temperaturas y humedad (INVIAS, 2013).



Figura 1. Pavimento Flexible. Tomado de Eadic, 2015.

Por otra parte, el pavimento semirrígido es similar al pavimento flexible, pero incorpora capas estabilizadas con materiales como cemento, cal o emulsiones asfálticas. Esto le permite combinar la capacidad de distribución de cargas propia del pavimento flexible con una mayor rigidez estructural. Suele usarse en corredores de alto tránsito, zonas industriales y patios de maniobra, asimismo tiene menor deformación permanente que el flexible y es susceptible a la aparición de fisuras por contracción.

Mientras que el pavimento rígido, compuesto por losas de concreto hidráulico o también llamado cemento portland, esto apoyado sobre una base granular o tratada, tiene como funcionamiento distribuir las cargas sobre un área amplia debió a la alta rigidez del concreto, esto para reducir esfuerzo en la subrasante. Sus principales características incluyen una alta durabilidad, con una vida útil aproximada de 20 a 30 años y la necesidad de un mantenimiento menos frecuente, aunque con mayores costos. Además, presenta una mayor resistencia a cargas pesadas, por lo que suele emplearse en autopistas con alto flujo vehicular, aeropuertos, puertos y vías que requieren una elevada capacidad estructural.



Figura 2. Pavimento Rígido. Tomado de Fripal, 2022.

Por último, está el pavimento articulado, conformado por adoquines de concreto o materiales similares, dispuestos sobre una cama de arena y confinados lateralmente, estos adoquines trabajan como una unidad que transmite las cargas por Inter trabado mecánico, es decir, que los adoquines trabajan en conjuntos para soportar el peso y no de forma aislada. Este tipo de pavimento presenta una alta resistencia superficial al desgaste, permite el reemplazo sencillo de las piezas dañadas y ofrece una buena estética urbana. Por estas razones, se utiliza comúnmente en vías locales, ciclorrutas, plazas, parques y patios de carga.

Así como se ha mencionado previamente, los pavimentos ya sean rígidos o flexibles distribuyen el peso entre sus diferentes capas. En ambos casos, la carga es recibida inicialmente por la capa superior y luego la presión se va dispersando hacia abajo. La base y la subbase ayudan a que el suelo natural o subrasante no reciba todo el peso de manera concentrada. Sin embargo, si las capas no están bien diseñadas, pueden presentarse fallas como: grietas por tracción en la parte inferior de la carpeta asfáltica, ahuellamientos cuando el suelo se deforma por exceso de presión o fisuras en las losas de concreto cuando estas no soportan adecuadamente los esfuerzos de flexión.

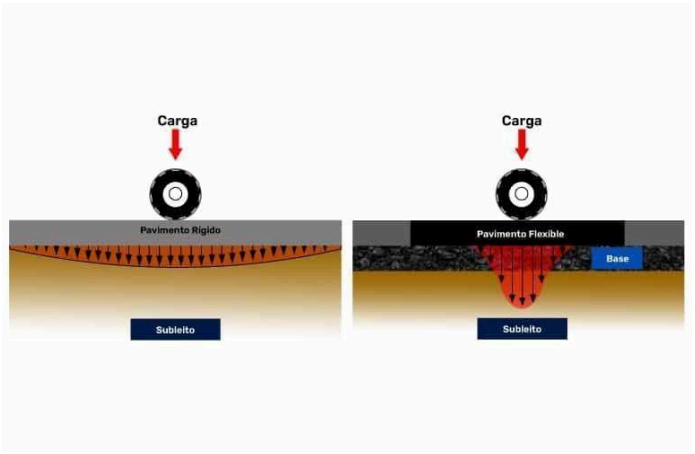


Figura 3. Esfuerzo pavimento flexible y rígido Tomado de Ekimak Engineering, 2023.

Una vez revisada la clasificación de los pavimentos, se profundizará en la estructura del pavimento flexible, que es uno de los más utilizados en varias ciudades principales, entre ellas Bogotá. Este tipo de pavimento se caracteriza por estar conformado por varias capas superpuestas, donde cada una cumple una función específica y se complementa con las demás. El buen desempeño de la vía depende de la adecuada interacción entre estas capas; sin embargo, un diseño inadecuado o el deterioro de alguna de ellas puede dar lugar a diferentes fallas. A continuación, se describen cada una de sus capas:

- **Carpeta Asfáltica:** Esta capa es donde va el pavimento caliente o concreto dependiendo de la necesidad, la capa superior y la que está más expuesta al tránsito y todos los cambios climáticos que haya. Cuando falla esta capa aparecen fisuras y baches.

- **Base Granular:** Es la capa intermedia, más resistente, fabricada con material pétreo triturado, esta capa está encargada de distribuir las cargas del tránsito hacia la subbase y aporta la rigidez a la estructura. Si hay una falla en esta capa, se presentan deformaciones permanentes.

- **Sub-base Granular:** Generalmente es construida con grava naturales o trituradas, arena y material filtrante, esta capa es más económica que la base y protege a la capa subrasante de la humedad y heladas, además de distribuir las cargas hacia la subrasante. Si esta capa falla se produce hundimiento o pérdida total de la capacidad estructural.

- **Subrasante:** Es el suelo de fundación, sobre el cual descansa el pavimento, esta capa se encarga de darle soporte a la estructura, garantiza una estabilidad volumétrica, que evita expansiones, retracciones y asentamientos. Si falla, puede ocasionar lo mismo que el fallo en la capa subbase.



Figura 4. Capas para la construcción de la vía Tomado de Giordani, s.f.

3.2. Mezcla Asfáltica

Las mezclas asfálticas son materiales compuestos resultado de combinaciones de agregados pétreos como gravas (partículas gruesas, generalmente trituras), arenas (Fina, natural o triturada), fillers o material de relleno (cal o cemento) con ligante, generalmente cemento asfáltico que sirve como un ligante que une los agregados. Tienen como función principal conformar la carpeta asfáltica de un pavimento, para brindar resistencia estructural, impermeabilidad y adherencia, pero la mezcla dependerá de factores como qué tanto flujo de tránsito hay o el clima:

La mezcla asfáltica en frío se produce y se aplica a temperatura ambiente, sin necesidad de calentar los agregados ni el ligante. Generalmente, se utiliza una emulsión asfáltica como ligante, la cual se mezcla con los agregados ya sea en planta o directamente en el lugar de la obra. Este tipo de mezcla se emplea principalmente en reparaciones rápidas y temporales, como el parcheo de vías, especialmente en zonas de bajo tránsito o áreas rurales. Entre sus ventajas se encuentran su bajo costo y la facilidad para su manipulación y almacenamiento; sin embargo, presenta una menor resistencia estructural en comparación con otros tipos de mezclas (INVIAS, 2012).

Las mezclas asfálticas tibias han ganado importancia en los últimos años. Se producen a temperaturas intermedias, generalmente entre 100 °C y 140 °C, mediante el uso de aditivos químicos, ceras o procesos de espumación con agua que reducen la viscosidad del asfalto y facilitan su manejo a menores temperaturas. Este tipo de mezcla se emplea tanto en carreteras urbanas como rurales con niveles de tránsito medio a alto, así como en proyectos donde se busca disminuir el impacto ambiental, ya que permite reducir las emisiones contaminantes y de gases de efecto invernadero. Además, ofrece una buena compactación incluso a temperaturas más bajas. Sin embargo, su implementación implica un mayor costo inicial debido al uso de aditivos y requiere personal con capacitación técnica adecuada para su correcta aplicación (Correa & Palacio, 2017).

Las mezclas asfálticas en caliente se producen a temperaturas que suelen oscilar entre 140 °C y 160 °C, utilizando cemento asfáltico convencional calentado para reducir su viscosidad. Este tipo de mezcla es ideal para vías de alto tránsito, como autopistas, avenidas principales, aeropuertos y terminales de carga, ya que ofrece una alta resistencia a esfuerzos repetitivos, buena durabilidad frente a cargas pesadas y una mayor impermeabilidad. Sin embargo, su producción implica un alto consumo energético y genera mayores emisiones contaminantes, como CO_2 y gases de combustión. Además, debido a que la mezcla se enfría rápidamente, es fundamental realizar su transporte y colocación de manera ágil para asegurar un buen desempeño (Correa & Palacio, 2017).

3.3. Pavimentos en Bogotá

La ciudad de Bogotá, al ser una de las más congestionadas del mundo, transporta diariamente en sus vías a miles de ciudadanos que se movilizan, en el 2024 el IDU indica que Bogotá siendo la capital de Colombia cuenta con 14,587.85 kilómetros-carril de malla vial, la cual se divide en urbana y rural, siendo la mayor proporción en la urbana, donde asimismo hay más habitantes, pero esta vía urbana en Bogotá, se divide en malla vial troncal, que es por donde pasan los buses del sistema de Transmilenio, teniendo en la ciudad carriles exclusivos para estos buses. Por otro lado, también está la malla vial arterial, donde hay un tráfico mixto y por último están las vías intermedias, por donde suelen haber rutas del transporte público de la ciudad.

En Bogotá el mantenimiento de la malla vial está encargado por dos grandes entidades el Instituto de Desarrollo Urbano (IDU) que se encarga de la malla vial troncal y arterial, y la Unidad de Mantenimiento Vial (UMV) responsable de la malla rural intermedia y rural. Estas dos entidades hacen una clasificación en siete categorías para el estado de la malla vial, los cuales son: Bueno, satisfactorio, justo, pobre, muy pobre, grave y fallado, teniendo el porcentaje más alto en la clasificación buena con el 38%, mientras que el menos presente, es la clasificación grave (IDU, 2023).

Asimismo, las entidades del IDU y la UMV, junto con la alcaldía de la ciudad han buscado aumentar la integración y ayuda por parte de todos los ciudadanos de Bogotá, es por esto que han creado diferentes alternativas para que los ciudadanos puedan reportar los baches de la capital, entre estas alternativas innovadoras, la UMV creó una página web / aplicación móvil, la cual consta de un mapa de toda la ciudad donde el ciudadano debe registrarse e ingresar fácilmente con la cuenta de Google, ingresar la dirección en la que se encuentra y llenar los datos donde tiene que poner el ancho y largo del bache, además de la fecha, observaciones e incluso fotografías, esto ayuda a tener un mayor seguimiento de los baches en la ciudad, pero esta idea innovadora no es tan

conocida por los ciudadanos y requiere de capacitaciones para muchos de los habitantes de la ciudad (Noticias Caracol, 2024).

Por otro lado, la Alcaldía de Bogotá, a través de su página web, habilitó un formulario sencillo para que los ciudadanos puedan reportar baches en la ciudad. Este formulario consta de dos secciones. La primera, denominada “Tus datos de ubicación y contacto”, solicita información básica del usuario, como nombre, correo electrónico y número de teléfono.

La segunda sección, llamada “Datos de tu reporte”, se enfoca en la caracterización del bache. Allí se pide indicar si se encuentra en una vía pública, el tiempo aproximado que lleva el daño, el nivel de tráfico en la zona, la dirección exacta y, de manera opcional, adjuntar una fotografía. Toda esta información es enviada directamente al Sistema Distrital de Quejas y Soluciones de Bogotá para su gestión (Alcaldía de Bogotá, 2023).

3.4. Fallas en el Pavimento

El desgaste o fallas en los pavimentos es una de las problemáticas más comunes, pero antes, es necesario entender cuál es la definición de una falla en el pavimento, este suceso es cualquier manifiesto visible o medible de deterioro que reduce la capacidad estructural o la funcionalidad de la capa de rodadura, tales como: Grietas, baches, deformaciones, desprendimiento de agregados, exudación del ligante, entre otros (U.S Department of transportation, 2003).

Los pavimentos no son duraderos para siempre, al contrario, requieren de un mantenimiento y planeación muy exhaustivos; Como se dijo anteriormente, una de las principales razones del desgaste de los pavimentos es la interacción que tiene la vía con vehículos de carga muy pesada, esto genera una fatiga y gran deformación en el pavimento.

De igual forma, la filtración de agua por climas muy lluviosos hace que se reduzca la rigidez de la subrasante. Por otra parte, también depende mucho del desgaste que ya tenga previamente el pavimento, como ya fue descrito anteriormente, el pavimento dependiendo de su

fabricación y el tipo que sea, tiene una vida útil aproximada, al pasar ese tiempo se genera un desgaste muy notorio en la vía, por eso es necesario rejuvenecer el pavimento viejo. De igual manera, en obras de hace muchos años, también pudo haber factores como el de hacer mezclas de baja calidad, que la compactación no sea suficiente, incluso que hayan realizado malas pruebas de calidad. Por lo que es necesario entender que el pavimento presenta distintas fallas, dependiendo de distintos factores, a continuación, se presentarán las fallas en las que se trabajarán:

Baches: Cavidades puntuales en la carpeta asfáltica; la superficie colapsa en zonas donde el soporte se ha perdido.



Figura 5. Bache en Bogotá Tomado de Sarmiento, 2012.

Fisuras/Grietas: Son deformaciones lineales o fracturas en la superficie del pavimento, causadas por factores como el envejecimiento del material, la repetición de cargas excesivas, los cambios térmicos, la saturación de agua, el asentamiento del terreno o de movimientos estructurales.



Figura 6. Fisura y Grieta Tomado de Ocampo, s.f.

Fatiga / deterioro estructural: Grietas que se multiplican hasta generar un patrón tipo "piel de cocodrilo" cuando la estructura está sobrecargada o la base/subbase no es la adecuada.



Figura 7. Piel de cocodrilo Tomado de Blogspot, 2014.

3.5. Técnicas de reparación

De la misma manera como hay fallas, existen bastantes alternativas para la reparación y rehabilitación en los pavimentos, una de las más famosas y utilizadas es el bacheo y parcheo. Esta técnica permite realizar un mantenimiento preventivo y focalizado, más específicamente se ponen

unos parches provisionales con mezclas frías, que son muy baratas y rápidas, pero duran mucho menos, mientras que los parcheos permanentes, duran mucho más.

En Bogotá, la Unidad de Mantenimiento Vial (UMV) emplea máquinas tapabaches para realizar intervenciones superficiales de rápida ejecución, lo cual evidencia que una proporción significativa de las reparaciones viales son de carácter temporal. Complementariamente, para el tratamiento de grietas y fisuras de menor tamaño, se recurre a la inyección de emulsión asfáltica. Esta técnica preventiva tiene como propósito fundamental sellar las aberturas para mitigar un deterioro mayor, impidiendo la infiltración de agua que acelera el daño de la capa asfáltica. Sin embargo, factores particulares de la ciudad, tales como el alto volumen de tráfico pesado, las variaciones climáticas constantes y las deficiencias en los sistemas de drenaje, disminuyen significativamente la eficacia y durabilidad de estas intervenciones.



Figura 8. **Procedimiento para Trabajos de Bacheo Tomado de Frapial, 2022.**



Figura 9. Sellador de fisuras, Tomado de Multinsa, 2022.

3.6. Trabajos Asociados

Se realizará una búsqueda y revisión de diversas investigaciones relacionadas con el tema, con el fin de analizar estudios previos, comprender cómo fueron desarrollados, qué problemáticas identificaron y cuáles fueron sus principales limitaciones u observaciones. Esto permitirá enriquecer el desarrollo del trabajo de grado.

En primer lugar, Ríos et al. (2020) presentan una recopilación y comparación de resultados de diferentes autores, basada en múltiples investigaciones. En este estudio se abordan técnicas de visión por computador, como la descomposición, el enfoque basado en semillas y los métodos de umbral.

Asimismo, los autores comparan estas técnicas con enfoques de aprendizaje automático, entre los que se incluyen las máquinas de soporte vectorial (SVM), los árboles de decisión, las redes neuronales y el aprendizaje profundo. Estas metodologías han sido evaluadas para la detección de fallas superficiales en pavimentos. Para ello, los distintos estudios analizados emplearon diversas técnicas de recolección de datos, como sistemas láser con triangulación, tiempo de vuelo y luz estructurada. También se destaca el uso de cámaras digitales para escaneo de línea, escaneo de área, video y análisis multiespectral, así como sistemas comerciales desarrollados por compañías internacionales que han equipado vehículos para la captura de información sobre el estado del pavimento. De manera complementaria, se utilizaron tecnologías como ultrasonido, acelerómetros, sensores infrarrojos y dispositivos RGB-D.

En esta revisión, se presenta una tabla que evidencia el porcentaje de uso de las técnicas de visión por computador y aprendizaje automático a lo largo del tiempo. Los resultados muestran que el aprendizaje automático ha incrementado significativamente su uso, pasando de un 9 % entre 1999 y 2008, a un 7 % entre 2009 y 2014, y alcanzando un pico del 41 % entre 2015 y 2019. Por su parte, la visión por computador tuvo un inicio del 17 % entre 1999 y 2008, alcanzó su punto más alto entre 2009 y 2014 y posteriormente descendió al 12 % en el periodo de 2015 a 2019.

En cuanto al desempeño de estas metodologías, la mayoría de los estudios reportaron resultados superiores al 80 % (0,8) en diferentes métricas, tales como Classification Accuracy Rate (CAR), F1-score, Root Mean Square Error (RMSE), puntuación de Hausdorff, correlación, tasa de error de reconocimiento y porcentaje promedio de error. No obstante, es importante señalar que el número de imágenes utilizadas en los estudios fue muy variable, desde conjuntos muy pequeños (4 imágenes) hasta otros mucho más amplios (2000 imágenes), lo que en algunos casos podría generar problemas de sobreajuste (overfitting) y explicar los altos valores obtenidos.

Adicionalmente, se analizó el tipo de fallas más estudiadas en la literatura. Las fisuras representan el 77 % de los casos analizados, seguidas por los baches (22 %), parches (6 %), ondulaciones (4 %), surcos (4 %), estimación de la condición global (4 %), marcas visuales (4 %) y, finalmente, las tapas de alcantarilla (2 %).

Por otro lado, en la Universidad de Antioquia, Escobar et al. (2023) presentan un estudio titulado “Estimación de irregularidades en pavimentos mediante técnicas de procesamiento digital de imágenes”. En este trabajo se desarrolla un algoritmo orientado a la identificación de tres tipos de fallas: baches, grietas y piel de cocodrilo, con el objetivo de estimar características como el área, el perímetro y el diámetro equivalente. Aunque los autores trabajaron con un conjunto reducido de solo cinco imágenes, resulta especialmente interesante el diagrama de flujo que proponen, ya que explica de manera clara el procesamiento aplicado a las imágenes y el funcionamiento del algoritmo.

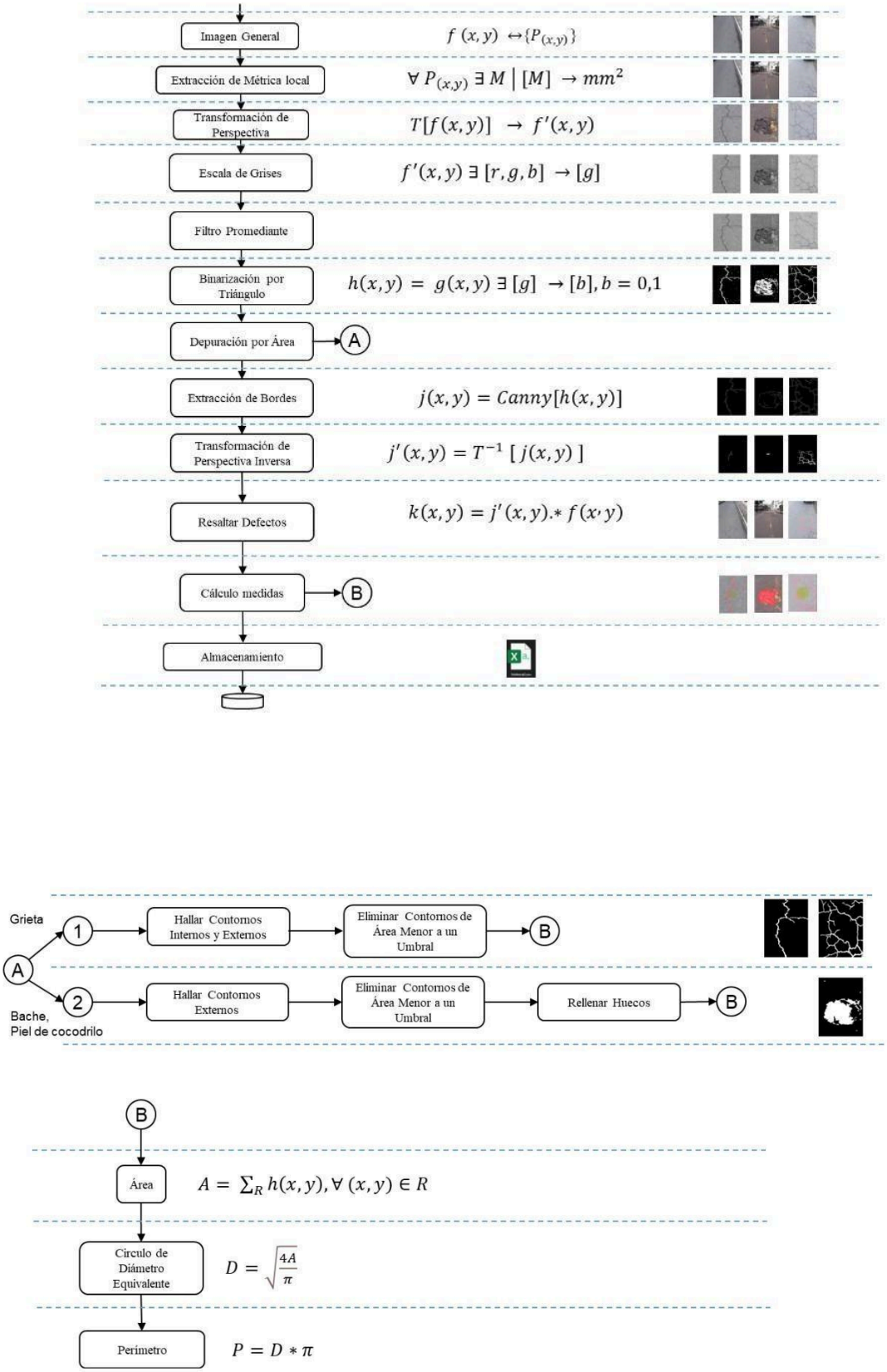


Figura 10. Diagrama de flujo del funcionamiento para el procesamiento del algoritmo (Escobar, 2023).

En sus resultados ellos solo los evaluaron mediante el concepto de la intersección sobre la unión, la precisión acá se obtiene, mediante la comparación de dos imágenes concernientes a una máscara real y una máscara de predicción, la cual en este caso será la identificación de irregularidades en pavimentos. En esta métrica se realiza la comparación entre la máscara directiva y la real, por medio de un algoritmo se determina un número del 0 al 1, siendo el 1 el valor óptimo y el que se esperaría tener, es por esto que matemáticamente la intersección sobre la unión (IoU) queda definida como:

IoU = Área de intersección de las dos imágenes / Área de unión de las dos imágenes

Con las 5 imágenes que se recolectaron se hizo este cálculo con el algoritmo y arrojó los siguientes resultados, juntos con la vista de las máscaras tanto reales como predictivas:

Imagen	IOU
1	0.69
2	0.87
3	0.79
4	0.87
5	0.81

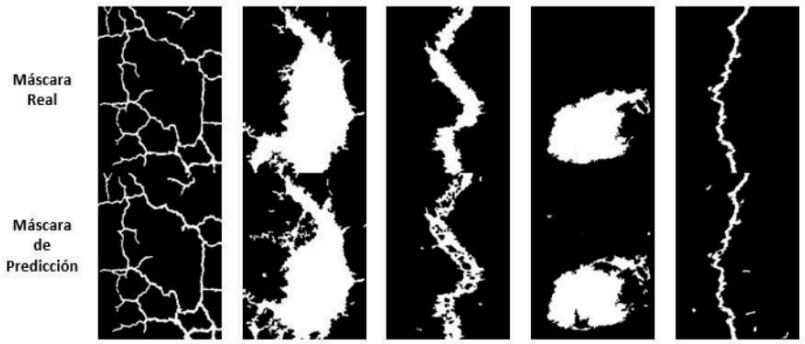


Figura 11. Máscaras y resultado IoU (Escobar, 2023)

Los autores comentan que la diferencia entre las máscaras con las etiquetas no está siendo tan precisa debido a el color más oscuro que generan las rocas pequeñas incrustadas en el pavimento o dentro de la misma imperfección.

Es por factores como la luz o sombra al momento de tomar la fotografía influyen en los resultados, indicando que las diversas transformaciones de perspectiva que se hacen pueden reformar los píxeles que hay dentro de la imagen, generando que las dos máscaras no coincidan en su totalidad. Aunque este artículo no especifica mucho lo que se hizo, es muy interesante y útil el procesamiento que usaron en las imágenes, además de la forma en sacar las métricas mediante IoU.

Posteriormente, en un trabajo de grado de la Pontificia Universidad Javeriana, donde Suarez (2023), muestra un "Sistema para analizar señales de sensores en celulares que permitan identificar y analizar la mala calidad del pavimento en las vías", en este trabajo de grado el autor diseñó una aplicación que permite obtener lo datos que genera los sensores de los celulares con el acelerómetro, GPS y giroscopio, esto para analizar la calidad del pavimento en la ciudad de Bogotá, los datos que son recogidos por el sensor del celular serán enviados a una base de datos para que posteriormente vaya a una página web, donde se generen dashboards con esa información recolectada.

Una vez el autor fue a realizar la calificación de las vías, usando los datos previamente recolectados, para así convertir los datos que generaba el sensor del celular en direcciones físicas reales, dividiéndola en 3 variables que formarán la dirección final. El cambio de las variables 1 y 2 son mandadas a una función asincrónica que se encargará de calcular la varianza, con esta varianza el autor crea una cualificación y una calificación que está basada en el rango numérico correspondiente a la varianza. El autor se encargó de clasificar el estado de la vía en tres: Buena con un rango de 4 a 5 correspondiente a una varianza entre 0 y 0.1202, regular con un indicador de 2 a 3.9, con una varianza entre 0.1203 a 0.5206 y por último está la calificación mala de la vía que tiene un indicador de 0 a 1.9 con varianzas entre 0.5207 y 2.1209.

Para la identificación de un bache con estos sensores, el automóvil debía pasar sobre el bache y se realizó una validación de la aceleración en el eje junto con el acelerómetro. Si esta

aceleración cae por debajo de -0.2 mientras el vehículo y el dispositivo móvil se desplazan hacia abajo, se genera un evento *(Suarez, 2023)*. Una vez detectada la anomalía, el sistema transforma los datos de latitud y longitud para saber dónde está el bache en direcciones físicas reales. Aunque este proyecto estaba más enfocado al internet de las cosas (IoT), dejó detalles muy interesantes como el uso de la varianza para sacar una métrica propia del estado de la vía, también fue interesante la forma en la que el autor captura los datos con unos sensores del teléfono.

En el paper titulado "Pavement crack detection algorithm based on image processing analysis", donde Qingbo (2016) hace un énfasis más matemático acerca del análisis de las imágenes para detectar un grieta o bache en el pavimento. Lo primero que hace el autor es modelar la imagen en una ecuación para corregir la iluminación o trasfondo, esto hace que la imagen esté normalizada para mejorar la segmentación que habrá posteriormente. La fórmula que propone el autor es al siguiente:

$$I(x, y) = Ib(x, y) + In(x, y) + Ic(x, y)$$

Donde:

$Ib(x, y)$: La iluminación Heterogénea

$In(x, y)$:Son las grietas

$Ic(x, y)$: Es el ruido

Posteriormente, el autor procede a reducir el ruido mediante un filtro de promedio ponderado por el recíproco del gradiente, en este concepto para cada píxel central en una ventana (n = 3), se debe calcular el gradiente absoluto entre el píxel central y cada vecino, tomando el

recíproco de ese gradiente como el peso, por lo que los pixeles que están dentro de la misma región tienen un peso alto lo que indica un gradiente pequeño, asimismo los pixeles que están cerca de un borde tendrán un peso bajo pero un gradiente grande.

Lo que continúa en el paper habla sobre que el nuevo valor del píxel será la suma ponderada de los pesos, fijando el elemento central, normalizando los recíprocos antes de hacer la matriz de pesos, esto lo que hace es eliminar el ruido aislado de la imagen preservando la información del borde. Posteriormente se hace una intensificación de imagen, esto mediante la coalición de histograma, para ampliar el rango dinámico, contrastar grietas frente al pavimento. Lo que hace la ecualización es aumentar el contraste, lo que puede generar que la imagen se vuelva borrosa o que los bordes se expongan de una forma desigual, por eso es importante hacer un buen filtrado.

Continuando se habla sobre los niveles de grises, la imagen ya debió ser llevada a escala de grises, por lo que, para la detección de bordes, lo que se usa es la derivada del gradiente, como el borde es una zona con cambio rápido de nivel de gris, se hacen estas variaciones a la derivada del gradiente, para esto el autor prueba 5 operados

a. **Roberts** (kernel 2×2): Muy simple, pero produce bordes gruesos y localización pobre.

b. **Prewitt** (3×3): Aproximación del gradiente, mejor localización que Roberts.

c. **Sobel** (3×3): Similar a Prewitt pero incorpora una suavización (Mayor robustez al ruido).

d. **Laplace**: Operador de segunda derivada; detecta *zero-crossings* y produce bordes completos y bien localizados, en el resultado del paper el autor afirma que Laplace fue "obviamente superior" en precisión y completitud de borde.

e. **Canny**: Multietapa — suavizado (Gaussiano), cálculo de gradiente, Non-Maximum Suppression (NMS) para afinar a un borde de un píxel, y hysteresis thresholding (Conexión por umbrales) — el paper concluye que Canny produce el borde más completo, debido a NMS y a la conexión morfológica posterior.

Ahora lo que prosigue es una segmentación del umbral, posteriormente a detectar los borde separa la "zona grieta" con el fondo, para esto recomienda la máxima varianza, se calcula la varianza intraclase para cada posible umbral *t* y se elige el *t* que maximiza esa varianza, así se puede separar la grieta y el pavimento. Continúa con la conexión y limpieza morfológica, donde recomienda usar operaciones morfológicas (dilatación, cierre, conexión) para unir bordes rotos y obtener una máscara coherente de grietas. También menciona la importancia de preservar la "completud del borde" (lo que favorece a Canny + morfología). Por último, el autor hace una extracción de características, una vez binarizadas y conectadas las grietas, extrae los siguientes atributos: longitud de trazas, ancho, orientación, patrón, densidad (longitud por área) y otras métricas texturales. Estos atributos sirven para clasificar la severidad y para calcular indicadores de condición.

A Continuación, se realizó una serie de pasos para que este proceso se entienda de una manera más simple y pueda ser captada la idea mucho mejor:

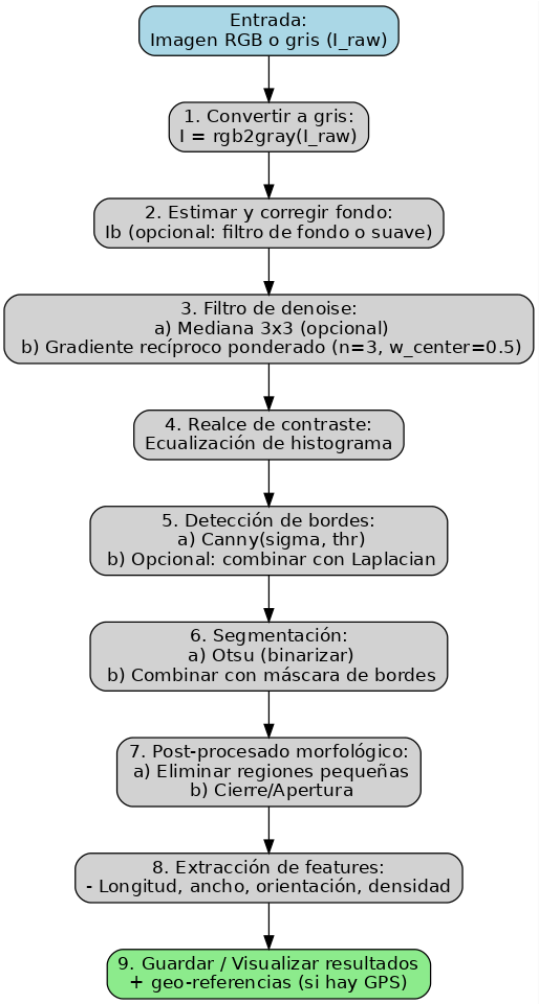


Figura 12. Diagrama proceso del paper (Elaboración propia, 2025)

Pasando con el artículo de Tello et al., mencionan acerca de la "Evaluación de daños en pavimento flexible usando fotogrametría terrestre y redes neuronales", donde hablan acerca de la metodología interesante que usaron, siendo esta: i. Captura de las imágenes desde un automóvil, ii. Preprocesamiento que permite eliminar la distorsión presente en la imagen y la mejora del contraste, iii. Segmentación mediante técnicas de detección de bordes y la transformada de Beamlet, iv. Extracción de características, v. Clasificación con redes neuronales artificiales (RNA) y vi. Evaluación del área de afectación.

Uniendo varias temáticas con el paper anterior, estos autores hicieron la recolección de las imágenes mediante una cámara digital GoPro de 11 Megapíxeles, un celular Samsung S3 y una

estructura de soporte que irá sujeta a un automóvil, para que así se vaya recolectando las imágenes, teniendo como resultado 365 capturas a horas de la tarde (3-5pm) a una velocidad del automóvil de 25km, con el objetivo de reducir el efecto de borrosidad y sombras en las imágenes.

La investigación buscaba identificar 3 tipos de deterioro en el pavimento: Grieta longitudinal en la cual se procesaron 30 imágenes, piel de cocodrilo en la que se emplearon 84 imágenes para procesar y por último los baches con 15 imágenes de procesamiento, además de eso se usaron 10 imágenes en validación y entrenamiento respectivamente para cada tipo de falla, menos en bache donde solo se usaron 5 en la validación.

Consecuente a esto, se hicieron los preprocesamientos de las imágenes, corrigiendo distorsiones del lente gran angular, pasaron las imágenes a escala de grises y mejoran el contraste para reducir sombras y evitar uniformidades, después hicieron la segmentación para la detección de los bordes, aplicaron la transformada Beamllet que es un método bastante avanzado para detectar líneas y bordes incluso en imágenes ruidosas, usando umbrales entre 0.5 y 0.7 según el tipo de falla.

Aparte complementaron con operaciones morfológicas de erosión y dilatación para limpiar el ruido que pudiera haber, rellenaron espacios vacíos y dejaron el borde del daño más definido, incluso pudieron esqueletizar los bordes para analizar la orientación, esto para el caso de grietas longitudinales.

Continuaron con la extracción de características, calculando los momentos invariantes geométricos, que no dependieran de traslación, rotación y escala, también concatenaron los primero cuatro momentos de cada imagen, resultando al final un vector de 12 características.

Finalizando hicieron una clasificación con Redes Neuronales Artificiales (RNA), usando

Perceptrón Multicapa (MLP) con configuración (12-12-3), entrenado con algoritmo Levenberg–Marquardt (backpropagation), dando como resultados que la etapa de evaluación de nivel de deterioro se halla el porcentaje de área afectada y de acuerdo con este se clasifica en deterioro ligero, medio o alto. Mientras que en la etapa de clasificación se obtuvo un porcentaje del 96%(accuracy), del 96.97%(precisión), del 93.33%(recall)y un índice Kappa de 0.936.

Buchinger et al., (2014) en la Revista Brasileira de Computação Aplicada, muestran un algoritmo que tiene dos fases principales. La primera fase es segmentar la vía y separarla de los demás objetos que hay en el entorno como: El cielo, árboles, edificios, entre otros. Para esto los autores usaron el algoritmo watershed condicionado que consiste en un método de segmentación muy potente que trata la imagen procesada como un mapa topográfico, pasando inicialmente la imagen a escala de grises, posteriormente se recorta para eliminar objetos y ahí aplican el algoritmo morfológico HMAX, para suavizar la imagen y controlar los picos más importantes de la imagen, es decir eliminando posibles ruidos que tenga la imagen. Por último, hacen la creación de “marcadores” predefinidos para indicarle al algoritmo Watershed qué es "carretera" y qué es "entorno" para mejorar la precisión.

En la segunda fase hacen la identificación de la falla (Fisuras y baches) mediante el nivel de oscuridad en la imagen, si una parte es muy oscura significa que se resalta la falla sobre el pavimento. Algo muy interesante fue que en lugar de usar métodos de umbralización tradicionales, los autores proponen un método basado en el concepto estadístico de intervalos de confianza (C.I.) y la desviación estándar del histograma de la imagen representando los píxeles de 0 a 255, con μ representando la media (brillo promedio), σ como la desviación estándar (qué tan dispersos están los valores) y n con la cantidad total de píxeles (tamaño de muestra), entonces se pueden definir zonas de interés en función de cuán lejos están los píxeles del promedio.

Por lo que en el proceso de detección se calcula el histograma de la imagen segmentada, ignorando el fondo blanco, así los autores asumen que el pico del histograma corresponde al color del pavimento normal y se calcula un umbral de segmentación, restando un múltiplo de la

desviación estándar definido previamente con el C.I. Finalizando con la aplicación de operaciones morfológicas para limpiar la imagen binarizada, además los autores usan clausura para conectar componentes cercanos, por ejemplo si una grieta aparece fragmentada por ruido o sombras, la clausura puede conectarla para que el sistema la detecte como una sola falla. También usan apertura por área esto para eliminar el ruido y componentes pequeños con regiones con área menor que el umbral.

En cuanto a recolección y preparación de los datos, los autores no usaron un dataset propio, sino que extrajeron las imágenes mediante la creación de un script usando la API de Google direction, recolectando sistemáticamente 229 imágenes a lo largo de una ruta tomando capturas panorámicas y las recortaron a un formato estándar (5,120 x 2,048 píxeles) para simular una cámara montada en un vehículo.

Los autores hicieron una clasificación manual de las imágenes en 4 categorías principales: Fisuras (63 imágenes), baches (13), carreteras normales (123) y carreteras con sombras (30). Ya en cuestión de resultado y métricas usaron las clásicas (accuracy, precision y recall), resaltando su principal limitación, los problemas con las sombras y la presencia de otros vehículos, generando muchos falsos positivos.

Escalona et al., (2019) en el instituto politécnico Nacional de México presentan un enfoque en deep learning para detectar fisuras abordando el problema no como una serie de pasos de procesamiento, sino como una tarea de segmentación semántica de extremo a extremo (end-to-end). La arquitectura UNet tiene como clave su estructura de codificador - decodificador. En el primer aporte de codificador (encoder) funciona una red convencional tradicional mediante capas de convolución y max-poling, donde se reduce el tamaño espacial de la imagen mientras aumenta el canal de características, siendo el objetivo principal capturar el contexto de las imágenes, es decir, que la red pueda aprender qué son las fisuras a diferentes escalas. Por el otro lado el decodificador (Decoder) realiza un proceso inverso, utiliza convoluciones hacia arriba con

up-convolution para aumentar gradualmente el tamaño de imagen y recuperar la información espacial.

Los autores usan conexiones de salto o Skip connection para que se conecten directamente las salidas de las capas del codificador con las capas correspondientes del decodificador. Esto permite que la red combine la información del contexto que brinda el encoder junto con la información de la localización precisa del decoder. Los autores usaron 3 variantes de la Unet para encontrar un balance entre el rendimiento y eficiencia: U-Net-A: Esta fue la arquitectura original y más profunda con 23 capas convolucionales, después en la U-Net-B se usaron 11 capas, siendo la versión intermedia y la última fue la U-Net-C, la versión más superficial con solo 7 capas.

En la recolección de los datos, los autores también usaron datos extraídos de datasets públicos siendo dos fuentes principales: Crack Forest Dataset, este dataset estuvo compuesto por 118 imágenes Red, Green and Blue (RGB) de 480x320 píxeles tomadas con un iPhone 5, los autores en este usaron 100 imágenes en train y el resto para test en los modelos. Mientras que en el otro AigleRN, era un dataset compuesto con 38 imágenes en escala de grises de 991 x 462 píxeles de pavimentos en Francia, pero este dataset fue utilizado solo para probar la generalización que tenía cada modelo entrando.

Los autores tomaron resultados ya obtenidos de un modelo CrackForest para tener un punto de referencia y poder comparar los resultados de sus modelos, siendo el mejor U-Net.B, que presentó el mejor rendimiento en general, logrando identificar fisuras delgadas y continuas de una forma muy clara, además fueron los más robustos hacia las sombras y manchas de aceite que presentaban las imágenes.

4. Metodología Propuesta

4.1. Marco de Gestión del Proyecto

Durante la metodología propuesta se hará una adaptación de la metodología ágil denominada como Scrum. Este método se basa en un enfoque iterativo e incremental donde se divide en ciclos de duración fija, denominados Sprints. La metodología Scrum no cuenta con pasos rígidos sino posee un conjunto de valores, roles, eventos y artefactos que permiten abordar los problemas de forma adaptativa, entregando valor de forma incremental y productiva; Este método es muy popular por su capacidad para ofrecer una gran flexibilidad, transparencia y ritmo de trabajo sostenible (Schwaber & Sutherland, 2020).

Dada la flexibilidad de Scrum para adaptarse fácilmente a cualquier proyecto, y por su modelo iterativo mediante Sprint, permite que se evalúen los avances paso a paso para que al llegar al producto final se logre el mejor y óptimo sistema. Asimismo, es fundamental hacer un gran Product Backlog y definir muy bien cada Sprint; Teniendo en cuenta eso, a continuación, se explicarán las adaptaciones que se hicieron del Scrum para este trabajo de grado:

Primero que todo se definieron los roles y responsabilidad que tendrá cada miembro, en este caso al ser un trabajo de grado individual, no hay muchos miembros en el equipo por lo que los roles quedan de la siguiente forma:

- **Product Owner:** El autor de este trabajo de grado, será el responsable de la visión del producto y se encargará de gestionar el product Backlog.

- **Development Team:** El autor también será el único encargado ya que es el único responsable de desarrollar las tareas técnicas de este trabajo de grado, también garantiza la calidad con la que el trabajo se entregue.

- **Scrum Master:** El Co-Tutor será el guía en este proceso de desarrollo, ayudará a facilitar el proceso y asegurará que el marco Scrum se aplique de una forma correcta, revisando los Sprint, haciendo correcciones y feedbacks. También ayudará con problemáticas en el desarrollo técnico del proyecto.

- **Stakeholders:** En este caso el tutor, ofrecerá la dirección de alto nivel y los requisitos clave desde la perspectiva de la ingeniería civil y los objetivos académicos. También estará al tanto de los Sprint y de cómo va el proyecto.

Una vez definidos los roles en esta metodología propuesta es importante mencionar que se definieron sprints cada 2 semanas, con reuniones al finalizar cada Sprint para poder realizar un feedback sobre cada entrega de los Sprint y ver posibles mejoras durante el proceso. El Product Backlog se establece en un Google Sheets compartido al Tutor y Co-Tutor para que le hagan el respectivo seguimiento al proceso. Por lo que a continuación se explicaran cómo estará compuesto ese Google Sheets; Además de poner por escrito los requerimientos funcionales y no funcionales como lo dictamina el Scrum. Pero es importante resaltar que estos no estarán explícitamente en el Google Sheets por temas de visualización.

En cuanto a las columnas que se establecieron en el Google Sheets se encuentra:

a. **ID**: Será el identificador de cada tarea, esencial para la gestión y la comunicación.

b. **Requisito:** Agrupa un conjunto de tareas bajo un objetivo o funcionalidad más grande. Responde al "¿Para qué?", dando contexto y evitando que el backlog se convierta en una lista de tareas técnicas aisladas.

c. **Tarea Específica:** Es la unidad de trabajo más pequeña y accionable. Responde al "¿Qué hay que hacer?", explica la tarea que se realizará.

d. **Estimación:** Mide el tamaño relativo de una tarea (combinando esfuerzo, complejidad e incertidumbre) usando la secuencia Fibonacci (1, 2, 3, 5, 8, 13...), según se indica por Scrum con el fin de tener una planificación realista.

e. **Sprint #:** Asigna la tarea a un ciclo de trabajo específico (ej. "Sprint 1", "Sprint 2"), ayuda a organizar el Sprint Backlog.

f. **Fecha Inicio y Fecha Finalización:** Establece el marco de tiempo planificado para la ejecución de la tarea ayudando a crear un mapa de ruta, también servirá como una alerta temprana si el proyecto se está desviando del cronograma general.

g. **Prioridad:** Ordena las tareas según su importancia para el éxito del proyecto (ej. Crítica, Alta, Media, Baja).

h. **Estado:** Rastrea el flujo de trabajo de la tarea en tiempo real, para ver cómo va cada una de las tareas.

Una vez claro todos los Ítems del Scrum adaptados a este proyecto de grado, se procede a presentar un diagrama de flujo que representa el flujo del trabajo.

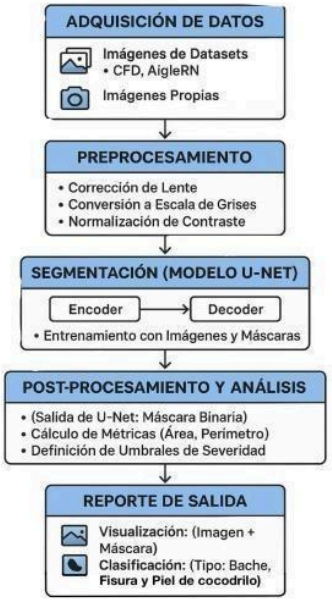


Figura 13. Diagrama de flujo proceso de desarrollo (Elaboración Propia)

Asimismo, se presentó el Product Backlog que se encontrará en el Google Sheets compartido a todos los miembros del equipo, además de mostrar un cronograma establecido basado en Sprint Backlog durante los dos semestres que dure el trabajo de grado.

Product Backlog — Semestre 1 | Sprints 1, 2 y 3 | Octubre – Diciembre 2025								
ID	Requisito	Tarea Especifica	Estimacion (Fibonacci)	Prioridad	Sprint #	Fecha Inicio (Planeada)	Fecha Final (Planeada)	Estado
A1	Fundamentacion	Documentar la metodologia Scrum y el marco teorico de la tesis (Cap. Metodologia).	3	Alta	1	13/10/2025	17/10/2025	Terminado
A2	Gestion de Datos	Descargar y organizar el dataset publico HuggingFace CrackSeg (7.732 imagenes).	3	Critica	1	18/10/2025	20/10/2025	Terminado
A3	Gestion de Datos	Realizar recoleccion propia de imagenes de asfalto en Bogota (minimo 50 fotos).	8	Alta	1	20/10/2025	25/10/2025	Terminado
A4	Desarrollo IA	Configurar el entorno de desarrollo en Google Colab (Python, TensorFlow, CUDA).	5	Critica	1	21/10/2025	23/10/2025	Terminado
A5	Gestion de Datos	Implementar script de preprocesamiento: escala de grises, redimension 256x256, normalizacion [0,1].	5	Critica	1	22/10/2025	28/10/2025	Terminado
A6	Gestion de Datos	Realizar etiquetado manual de mascaras (canal verde fisuras, canal rojo baches) en 20-30 imagenes.	8	Critica	1	27/10/2025	31/10/2025	Terminado
A7	Gestion de Datos	Dividir el dataset (publico + propio) en carpetas train / val / test (80-10-10).	1	Alta	1	31/10/2025	31/10/2025	Terminado
A8	Desarrollo IA	Codificar la arquitectura U-Net multiclase base (conv_block, encoder, bottleneck, decoder).	8	Critica	2	03/11/2025	04/11/2025	Terminado
A9	Desarrollo IA	Implementar el pipeline de entrenamiento con Data Loader tf.data, callbacks ModelCheckpoint y ReduceLROnPlateau.	13	Critica	2	04/11/2025	05/11/2025	Terminado
A10	Desarrollo IA	Ejecutar primer ciclo de entrenamiento completo (30 epocas, 3340 imagenes) y guardar mejor modelo.	8	Critica	2	05/11/2025	06/11/2025	Terminado
A11	Evaluacion	Generar mascaras de prediccion en el set de prueba y calcular metricas IoU, Dice y Recall.	3	Alta	2	06/11/2025	06/11/2025	Terminado
A12	Evaluacion	Crear visualizaciones comparativas (imagen original / mascara real / mascara predicha) y entregar resultados preliminares.	5	Alta	2	07/11/2025	07/11/2025	Terminado
A13	Documentacion	Redactar seccion 5 de la tesis: Resultados Preliminares y analisis de fallos del modelo unificado.	5	Alta	3	17/11/2025	28/11/2025	Terminado
A14	Fundamentacion	Realizar revision bibliografica de Transfer Learning, Attention U-Net y funciones de perdida avanzadas.	3	Media	3	01/12/2025	05/12/2025	Terminado
A15	Gestion de Datos	Generar metadata_clasificado.csv separando baches y fisuras por codigo de color de mascara.	3	Alta	3	08/12/2025	10/12/2025	Terminado
A16	Desarrollo IA	Disenar la arquitectura del sistema Multi-Experto (dos U-Net especializadas + fusion de predicciones).	5	Critica	3	10/12/2025	17/12/2025	Terminado
A17	Documentacion	Redactar plan detallado del Semestre 2: estrategias por especialista, hipotesis H0/H1 y metricas objetivo.	3	Media	3	15/12/2025	19/12/2025	Terminado

Leyenda de estados:	Terminado	En Progreso	Por Hacer	
Leyenda de prioridades:	Critica	Alta	Media	Baja

Figura 14. Producto Backlog propuesto para el primer semestre del trabajo de grado (Elaboración propia)

Product Backlog — Semestre 2 | Sprints 4 al 7 | Enero – Abril 2026								
ID	Requisito	Tarea Especifica	Estimacion (Fibonacci)	Prioridad	Sprint #	Fecha Inicio (Planeada)	Fecha Final (Planeada)	Estado
B1	Desarrollo IA	Integrar encoder ResNet34 pre-entrenado (ImageNet) como backbone de la U-Net para fisuras.	13	Critica	4	12/01/2026	16/01/2026	Terminado
B2	Desarrollo IA	Implementar Attention Gates en el decoder para suprimir fondo de asfalto y enfocar zonas de falla.	8	Critica	4	14/01/2026	20/01/2026	Terminado
B3	Desarrollo IA	Implementar perdida combinada Focal Loss (gamma=2, alpha=0.75) + Tversky Loss (alpha=0.7) para fisuras.	5	Critica	4	19/01/2026	22/01/2026	Terminado
B4	Desarrollo IA	Implementar Data Augmentation avanzado: 7 transformaciones geometricas y fotometricas sincronizadas imagen-mascara.	8	Alta	4	20/01/2026	26/01/2026	Terminado
B5	Desarrollo IA	Pre-entrenar especialista en fisuras con dataset HuggingFace (7.732 imgs, encoder congelado, 20 epocas).	8	Critica	4	26/01/2026	04/02/2026	Terminado
B6	Desarrollo IA	Fine-tuning del especialista en fisuras con dataset local de asfalto (encoder descongelado, LR=5e-5, 25 epocas).	13	Critica	4	02/02/2026	13/02/2026	Terminado
B7	Desarrollo IA	Construir U-Net con encoder VGG16 + Attention Gates y perdida BCE+Dice para especialista en baches.	8	Critica	5	09/02/2026	17/02/2026	Terminado
B8	Gestion de Datos	Calcular densidad de bache por imagen y definir estratificacion en 3 niveles para los folds.	3	Alta	5	16/02/2026	18/02/2026	Terminado
B9	Desarrollo IA	Implementar validacion cruzada estratificada de 5 folds (StratifiedKFold) para estimacion robusta con ~710 imagenes.	8	Critica	5	18/02/2026	25/02/2026	Terminado
B10	Desarrollo IA	Ejecutar loop de 5 folds: fase A encoder congelado (12 epocas) + fase B fine-tuning global (18 epocas) por fold.	13	Critica	5	23/02/2026	11/03/2026	Terminado
B11	Desarrollo IA	Implementar cosine LR schedule con warmup lineal (3 epocas) para ambos especialistas.	5	Alta	5	02/03/2026	05/03/2026	Terminado
B12	Desarrollo IA	Entrenar modelo final de baches con todos los datos train+val (fase A + fase B) usando mejor configuracion de folds.	8	Critica	5	09/03/2026	18/03/2026	Terminado
B13	Evaluacion	Evaluar ambos modelos en test set: calcular IoU, Dice, Precision, Recall y F1 por clase.	5	Critica	6	18/03/2026	21/03/2026	Terminado
B14	Evaluacion	Implementar busqueda automatica de umbral optimo de decision (rango 0.10-0.90) sobre conjunto de validacion.	3	Alta	6	19/03/2026	21/03/2026	Terminado
B15	Evaluacion	Implementar Test-Time Augmentation (TTA, 6 orientaciones) y medir mejora de IoU respecto a prediccion directa.	5	Alta	6	23/03/2026	26/03/2026	Terminado
B16	Evaluacion	Construir funcion diagnosticar_imagen(): overlay visual, mapa de confianza, clasificacion de severidad (ligera/moderada/severa).	8	Alta	6	24/03/2026	28/03/2026	Terminado
B17	Evaluacion	Generar tabla comparativa de resultados version anterior vs. version mejorada	3	Alta	6	27/03/2026	28/03/2026	Terminado
B18	Documentacion	Redactar seccion 6 de la tesis: Resultados Finales, evolucion arquitectonica y validacion visual cualitativa.	8	Alta	7	30/03/2026	08/04/2026	Terminado
B19	Documentacion	Redactar analisis de hipotesis H0/H1 (impacto de la perdida IoU) y discusion de resultados cuantitativos.	5	Alta	7	07/04/2026	11/04/2026	Terminado
B20	Documentacion	Completar bibliografias en formato APA 7, corregir gramatica, redaccion y citas en todo el documento.	5	Alta	7	10/04/2026	16/04/2026	Terminado
B21	Evaluacion	Realizar pruebas finales del sistema con imagenes propias de Bogota y documentar casos de exito y limitaciones.	5	Alta	7	13/04/2026	17/04/2026	Terminado
B22	Documentacion	Revision integral del documento con tutor y co-tutor, incorporar correcciones y preparar entrega final.	8	Critica	7	17/04/2026	28/04/2026	Terminado
B23	Documentacion	Entrega definitiva del trabajo de grado a la Pontificia Universidad Javeriana.	1	Critica	7	30/04/2026	30/04/2026	Terminado

Leyenda de estados:	Terminado	En Progreso	Por Hacer	Entregado
Leyenda de prioridades:	Critica	Alta	Media	Baja

Figura 15. Producto Backlog propuesto para el segundo semestre del trabajo de grado (Elaboración propia)

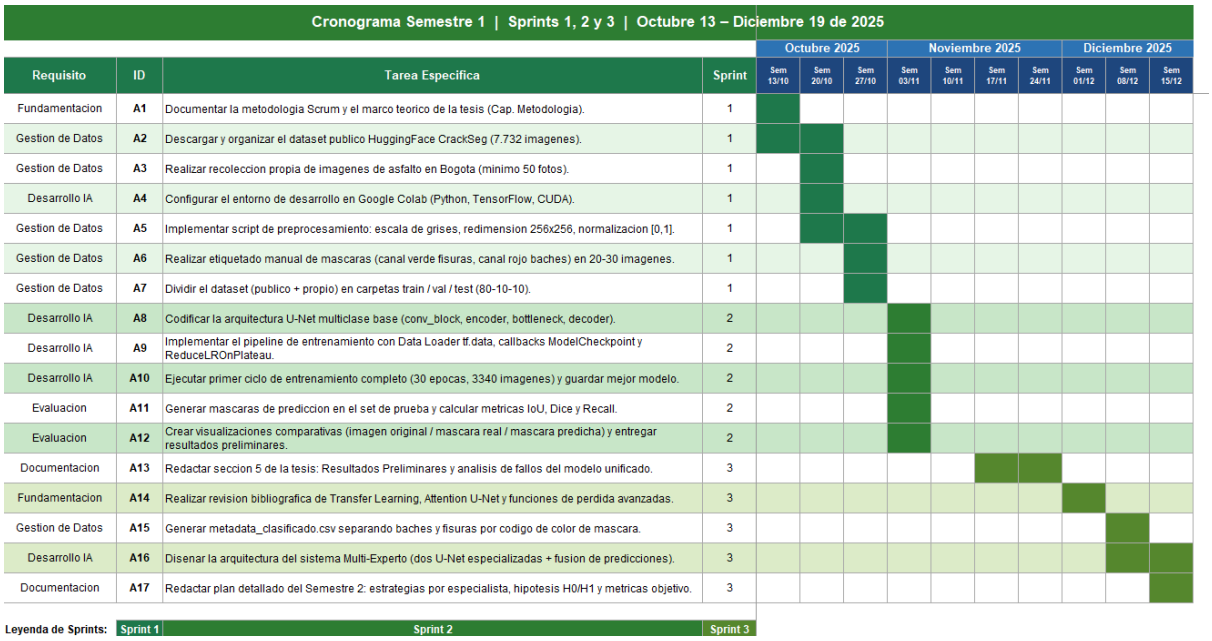


Figura 16. Cronograma propuesto para el primer semestre del trabajo de grado (Elaboración propia).

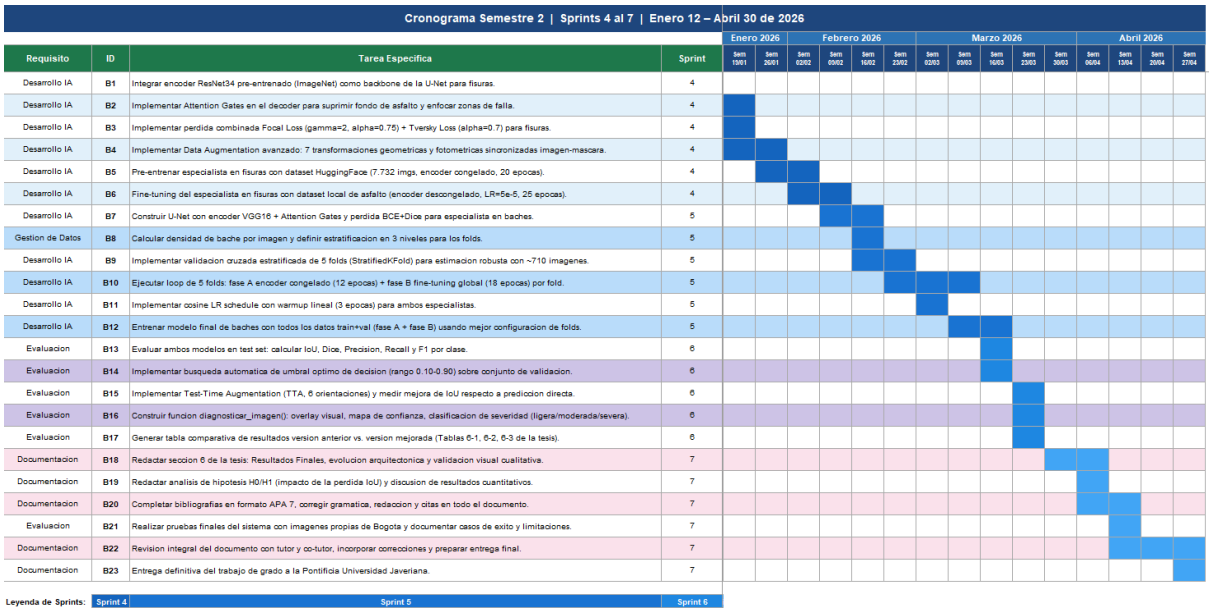


Figura 17. Cronograma propuesto para el segundo semestre del trabajo de grado (Elaboración propia).

4.2. Diseño del Sistema Multi-Experto y origen de datos

El problema de segmentación de fallas viales presenta una dificultad estructural que condiciona todas las decisiones de diseño del sistema: las dos clases de daño objetivo, fisuras y baches, tienen morfologías fundamentalmente distintas. Las fisuras son estructuras lineales delgadas y muy finas que en la mayoría de imágenes ocupan menos del 5% del área total, generando un desbalance extremo de píxeles frente al pavimento sano(Porras et al., 2012). Los baches, en contraste, son cavidades compactas de mayor área superficial con bordes irregulares pero definidos. Forzar a una única red neuronal a aprender simultáneamente ambos patrones deriva en confusión de clases, donde el modelo tiende a colapsar hacia la clase mayoritaria (el pavimento sano) e ignorar las fallas de menor representación (Rique et al.,2025).

Por esta razón se adoptó un enfoque Multi-Experto, en el que dos modelos de segmentación independientes se entrenan de forma especializada sobre su clase objetivo y posteriormente fusionan sus predicciones para generar un diagnóstico conjunto. Este enfoque permite adaptar de forma independiente la arquitectura, la función de pérdida, la estrategia de datos y los hiperparámetros de cada modelo a las características morfológicas específicas del daño que debe detectar.

La construcción del dataset de entrenamiento requirió integrar múltiples fuentes debido a la escasez de bases de datos públicas que combinarán imágenes de pavimento con máscaras de segmentación a nivel de píxel en formato compatible con la arquitectura U-Net. La literatura actual tiende a distribuir anotaciones de baches en formato de coordenadas de bounding boxes tipo YOLO, que encierran el daño en rectángulos y pierden la curvatura y el detalle a nivel de píxel que exige la segmentación semántica. Esta incompatibilidad obligó a descartar varios datasets externos y a construir la base de datos a partir de las siguientes fuentes:

Dataset principal local: obtenido a partir del repositorio compartido por el autor Farzad Nekouei en su trabajo de detección de baches con U-Net. Este dataset contenía 3340 imágenes organizadas en carpetas de train, test y valid, cada una con su máscara correspondiente. Para este

trabajo se utilizó únicamente la partición de train, correspondiente a las 3340 pares imagen-máscara, unificando todas las subcarpetas en dos directorios: imágenes y máscaras.

Además se usó un dataset HuggingFace CrackSeg: utilizado exclusivamente para el pre-entrenamiento del especialista en fisuras. Este contiene 7.732 pares imagen-máscara de fisuras en distintos materiales, incluyendo concreto, madera y pavimento. Dado que este dataset incluye fisuras en superficies distintas al asfalto, se empleó como fuente de conocimiento general sobre la morfología lineal de las fisuras, con posterior fine-tuning sobre el dataset local para la adaptación de dominio.

eRespecto a la categoría de baches, se consolidó un conjunto final de 568 imágenes con sus respectivas máscaras de segmentación. En contraste, para la clase de fisuras, se integraron 2.772 imágenes del conjunto de datos inicial con 7.732 muestras adicionales extraídas del dataset *CrackSeg* (HuggingFace), alcanzando un total de 10.504 imágenes y máscaras para fisuras. Adicionalmente, se implementó un script en Python para realizar un muestreo de los datos, con el propósito de identificar y estimar la distribución de los ángulos de captura de las fotografías en los datasets. El script de Python aplicó un muestreo estratificado. Esto significa que el código dividió el universo de datos en dos estratos (Clase: baches y clase: fisuras), también extrajo de cada división la cantidad exacta de imágenes necesarias para que la muestra mantuviera la misma proporcionalidad matemática que los datasets originales.

Para distinguir los ángulos de captura del conjunto de datos consolidado ($N = 11.072$), se optó por una inspección manual detallada sobre una muestra aleatoria estratificada de $n = 100$ imágenes. En estadística descriptiva, este tamaño muestral permite estimar las proporciones categóricas con un nivel de confianza del 95% y un margen de error cercano al ±9.8%. Respecto a la revisión cualitativa manual, esta se priorizó para garantizar un 'Ground Truth' exacto en la clasificación geométrica, evitando los sesgos que podrían proporcionar algunos algoritmos heurísticos en detección de bordes. En cuanto a la clasificación se tuvieron en cuenta 4 perspectivas (Cenital, frontal lateral y rasante), ya que estas cuatro perspectivas responden directamente a los distintos desafíos geométricos que enfrenta

una red neuronal convolucional, como la arquitectura U-Net, al momento de extraer características espaciales. Cabe resaltar que esta clasificación se hizo de forma manual, el script de python generaba la muestra y debido a esto, se fueron buscando cada una de las imágenes junto con su clasificación mediante un archivo csv que contiene la ruta de las imágenes de la muestra, si es bache o es fisura y por último, la columna que indica el tipo de ángulo que tenía esa imágen.

El ángulo cenital ocurre cuando la cámara captura el pavimento desde arriba, de forma completamente perpendicular al suelo. Es el escenario geométrico ideal, ya que ofrece una proyección ortogonal plana sin líneas de fuga.

La vista frontal corresponde a la perspectiva natural de un conductor o una cámara montada en un vehículo, apuntando a lo largo de la vía hacia el horizonte. Esta categoría es fundamental porque introduce variaciones drásticas de escala geométrica.

La captura lateral se presenta cuando la fotografía se toma en diagonal respecto a la vía, generalmente desde un andén o cruzando la calle. Este ángulo genera una distorsión asimétrica, donde el borde del daño más cercano a la lente retiene alta densidad de detalles, mientras que el opuesto sufre compresión visual.

Finalmente, el ángulo rasante agrupa las imágenes tomadas casi a ras de suelo o en un primer plano extremo para evidenciar hundimientos. En esta vista se pierde por completo el contexto global de la carretera y domina la oclusión topográfica, obligando al modelo a depender exclusivamente de texturas rugosas, sombras y cambios de iluminación para detectar el daño.

Los resultados fueron los siguientes:

Perspectiva de Captura	Cantidad (n=100)	Proporción Estimada
Cenital (Por encima)	72	72%
Frontal (Nivel de calle)	20	20%
Rasante (Primer plano)	5	5%
Lateral (Diagonal)	3	3%

Tabla 1. Resultados muestreo y clasificación de las imágenes. (Elaboración propia, 2026).

Los resultados de la muestra realizada indican una clara predominancia de la vista cenital (72%). Esta distribución resulta altamente favorable para el entrenamiento de arquitecturas convolucionales orientadas a la segmentación semántica, como U-Net. Al ubicarse la cámara de forma perpendicular al plano del pavimento, se reduce considerablemente la distorsión geométrica y de perspectiva, lo que permite que las máscaras de segmentación representen con mayor precisión el área real de la falla.

Por otra parte, la presencia de un 20% de imágenes con vista frontal contribuye a que el modelo adquiera mayor robustez y capacidad de generalización frente a escenarios reales de captura desde vehículos en movimiento, donde suelen presentarse líneas de fuga y variaciones en la escala geométrica.

No obstante, se evidencia una falencia en cuanto a la diversidad de perspectivas fotográficas disponibles, lo que podría limitar el desempeño del modelo dependiendo del ángulo desde el cual se capture la imagen. En este sentido, las perspectivas rasante y lateral, que en conjunto representan apenas el 8% de la muestra, constituyen un desafío para la generalización. Este aspecto podría abordarse en trabajos futuros o mediante esquemas multiexperto, incorporando técnicas específicas de aumento de datos geométricos, como transformaciones de perspectiva y cizallado, con el fin de balancear artificialmente estas vistas.

5. Resultados Preliminares

Para la automatización del diagnóstico de fallas en el pavimento, esta investigación implementa como motor principal la arquitectura de red neuronal convolucional conocida como U-Net. A diferencia de los modelos tradicionales de inteligencia artificial que se limitan a clasificar una imagen entera (indicando simplemente "esta foto contiene una vía en mal estado"), U-Net fue diseñada específicamente para una tarea mucho más compleja y topográfica: la segmentación semántica.

La elección de esta arquitectura arquitectónica sobre otros algoritmos populares de detección de objetos (como los modelos YOLO) radica en el nivel de detalle que exige esta problemática de la ingeniería civil. Los modelos tradicionales encierran los objetos encontrados dentro de cajas delimitadoras cuadradas (*bounding boxes*). Si bien esto es útil para contar vehículos o peatones, resulta insuficiente para evaluar pavimentos. U-Net, por el contrario, no dibuja una caja general; evalúa matemáticamente cada píxel individual de la fotografía para decidir si pertenece al pavimento sano o a una patología estructural de manera detallada.

En términos prácticos, esto significa que el modelo realiza un levantamiento perimetral exacto de la falla. Al delinear los contornos irregulares de un bache o el recorrido sinuoso de una fisura, U-Net genera un polígono preciso del área de afectación. Esta capacidad de medir la geometría real del daño es el insumo fundamental para que las entidades encargadas del mantenimiento vial puedan calcular áreas de parcheo, estimar volúmenes de material (mezcla asfáltica), priorizar el presupuesto y determinar la severidad exacta de la falla estructural sin depender de la subjetividad del ojo humano.

El nombre "U-Net" proviene de su diseño en forma de "U", el cual imita el proceso de inspección visual: una primera fase de "contracción" donde el modelo extrae el contexto general de la escena (luces, sombras y texturas globales del pavimento), y una fase de "expansión" donde utiliza

ese contexto para trazar topográficamente los bordes exactos del daño en su resolución original. Esta dualidad la convierte en la herramienta óptima para el diagnóstico automatizado de infraestructura.

5.1. Configuración Experimental y Adquisición de Datos

En este trabajo de grado se usó únicamente Python como lenguaje de programación. Así que A continuación, se describen las librerías implementadas y su función dentro del entorno de desarrollo.

- **Tensorflow / keras / models / layers / ModelCheckpoint**: Se utilizó para construir, compilar y entrenar la red U-Net, además de guardar el mejor modelo.

- **Pandas:** Se usó para leer y organizar el metadata.csv que conecta cada imagen con su máscara.

- **Numpy:** Se encargó de operaciones numéricas rápidas con los píxeles de las imágenes y máscaras.

- **Matplotlib.pyplot:** Utilizado para visualizar y graficar las imágenes de entrada y las máscaras que predicen el modelo.

- **Os:** Ayudo para manejar rutas de archivos y carpetas (como unir /content/gdrive/ con /Fallas/ en el drive).

- **Cv2 (OpenCV):** Usada para funciones específicas de procesamiento de imágenes, como para leer los valores de píxel.

Por otro lado, para este trabajo de grado se planteó en un inicio el uso de datasets que se hayan usado previamente en otros proyectos relacionados, pero casi ninguno de los proyectos, contaban con un acceso fácil a los datos. Por lo que se optó por buscar uno propio; Es así que mediante un video del autor Farzad Nekouei, titulado "Image Segmentation | Pothole Detection using U-Net and ResNet34". En este video el autor explicó el proceso para hacer las máscaras, aunque no fue muy detallista con el proceso, ni con las imágenes que se encontraban en el drive, el dejó en la descripción del video el link al drive donde se almacenaba el dataset.

Se logró encontrar un dataset bastante rico en imágenes de diversas fisuras y baches que era el objetivo principal, además de que en muchos datasets previamente buscados, había diversos problemas como que la cantidad de imágenes era muy baja o no contaban con las máscaras. Todas las imágenes estaban contenidas en un. Zip que estaba dividido en: Train, test y valid.

En este trabajo de grado, para el entrenamiento del modelo U-Net se utilizó únicamente el conjunto de train, el cual estaba conformado por 3340 imágenes a color con diferentes resoluciones en píxeles. Cada imagen contaba con su respectiva máscara. Inicialmente, los datos se encontraban organizados en varias carpetas; sin embargo, se unificaron y se cargaron en Google Drive, separando únicamente las imágenes y las máscaras en carpetas distintas.

5.2. Resultados del Preprocesamiento y Etiquetado

5.2.1. Pipeline de Preprocesamiento de Imágenes

Antes de poder entrenar la arquitectura U-Net, fue fundamental implementar un pipeline de datos robusto para procesar y limpiar el conjunto de 3340 imágenes y máscaras. Este proceso aseguró que los datos de entrada fueran uniformes, limpios y estuvieran correctamente etiquetados.

Ya en el Pipeline, cada imagen que iba entrando, pasaba por la siguiente serie de transformaciones:

Conversión a Escala de grises: Se hizo una conversión de todas las imágenes a escala de grises principalmente porque tiene una menor complejidad computacional. Al tratarse de fallas en el pavimento diferenciar colores no es algo muy relevante. Mientras que en cuestiones de sombras, polvo o iluminación puede generar un ruido innecesario en las imágenes. Además, esto ayuda a generalizar mejor, ya que las fotos fueron tomadas en distintas ciudades, en distintos puntos de iluminación y así el modelo se enfoque en las formas y patrones de daño, lo que mejora su desempeño en nuevos escenarios.

Redimensionamiento: Se ajustaron todas las imágenes a un formato de 256 x 256 píxeles ya que la arquitectura U-Net requiere un tamaño de entrada fijo para poder conectar las capas del codificador y el decodificador.

Normalización: Los valores de los píxeles que en un inicio iban de 0 a 255 se normalizaron para quedar en un rango de 0 a 1 dividiendo cada pixel por 255. Esto es fundamental para estabilizar el entrenamiento ya que el rango de los píxeles es más pequeño, y asimismo mejora la precisión numérica de la arquitectura U-Net.

5.2.1. Limpieza y Definición de Etiquetas

Esta fase fue muy importante ya que acá se iban a definir las categorías y fue muy importante analizar las máscaras que venían incluidas en el dataset; Por lo que se hizo la siguiente codificación:

Codificación de Clases: Las máscaras que venían en el dataset no usaban índices de clase simples como enumeración de 0 a 3, sino que codificaban las fallas usando valores de intensidad de gris por ejemplo 76 o 149. Así que se procedió a mirar las imágenes manualmente para seleccionar cinco que fueran baches y otras cinco que fueran agrietamientos. Esto con el fin de seleccionar el nombre de cada imagen, procesarla y ver como estaba clasificada, dando como resultado que constantemente los baches estaban en la escala de grises número 76. Mientras que los agrietamientos estaban con el valor de 149 en la escala de grises. Dando como resultado final la siguiente codificación de las clases:

- **Clase 0:** Fondo (Pavimento, carros, sombras, etc.)

- **Clase 1:** Bache (Mapeado desde el valor 76)

- **Clase 2:** Agrietamiento (Mapeado desde el valor 149)



Figura 18. Transformación a Escala de grises de las imágenes (Elaboración propia, 2025).



Figura 19. Dimensionalidad de las imágenes (Elaboración propia, 2025).

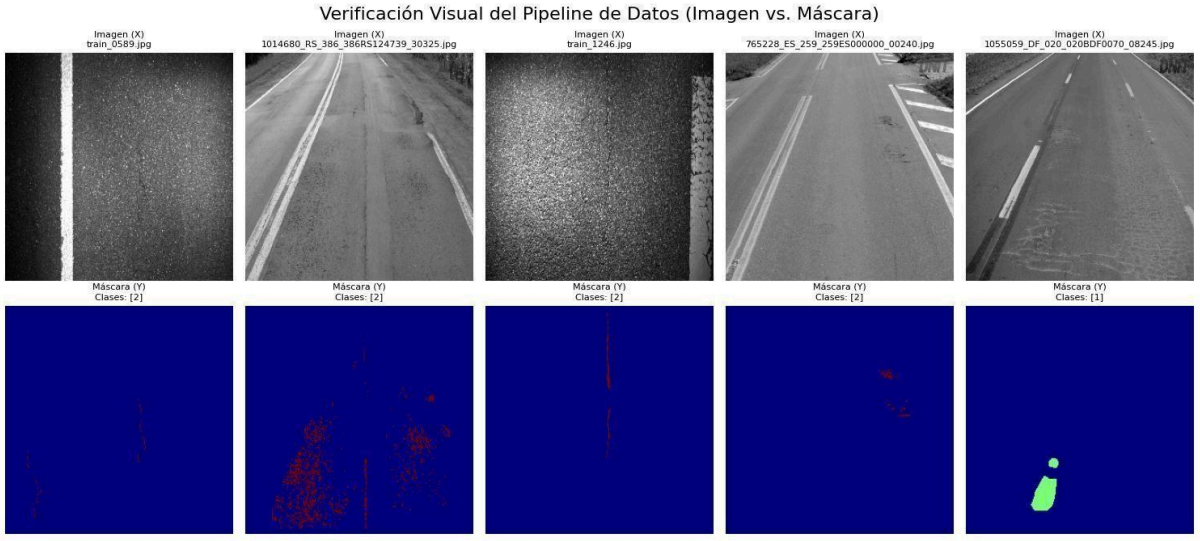


Figura 20. Visual de la imagen en escala gris con su categoría y respectiva mascara (Elaboración propia, 2025).

5.3. Primeros Resultados preliminares

Para validar el rendimiento del modelo entrenado, se optó por utilizar el conjunto de datos de prueba, el cual contiene 334 imágenes. Donde se hizo la siguiente serie de pasos:

Primero, se cargó el mejor modelo que se guardó anteriormente desde Google Drive. Este no representa el final de las 30 épocas, sino el modelo de la Época 23, que fue el que obtuvo el menor error de validación (val_loss: 0.05796) y, por lo tanto, es la versión más precisa. Luego se tomó un lote de imágenes de prueba y se pasaron al modelo. Pero el modelo no genera la máscara directamente. Su salida es un tensor con forma (lote, 256, 256, 3). El 3 representa las probabilidades para cada píxel de pertenecer a nuestras tres clases (0=Fondo, 1=Bache, 2=Agrietamiento). Así que se aplicó la función np.argmax a la salida. Esta función revisa las 3 probabilidades para cada píxel y "colapsa" el resultado, seleccionando la clase con el puntaje más alto. Este resultado es la máscara final, así que se graficaron las tres imágenes (Original, Máscara Real y Máscara Predicha) una al lado de la otra para una comparación visual directa.

Para comprender estso resultados preliminares, las siguientes figuras presentan una matriz comparativa que simula el debujo de la falla identificada. La visualización se divide en tres niveles horizontales:

- **Fila Superior (Imagen X):** Representa la fotografía cruda capturada desde el dataset.

- **Fila Central (Máscara Real - Y):** Es el dibujo de la falla que se tomo como referencia. Aquí se mapea topográficamente, donde el fondo azul oscuro representa el pavimento sano, el color rojo delimita los daños estructurales finos (fisuras) y el color verde demarca las zonas de pérdida severa de material (baches).

- **Fila Inferior (Máscara Predicha):** Muestra el diagnóstico generado por la primera versión del modelo U-NET (modelo multiclase unificado).

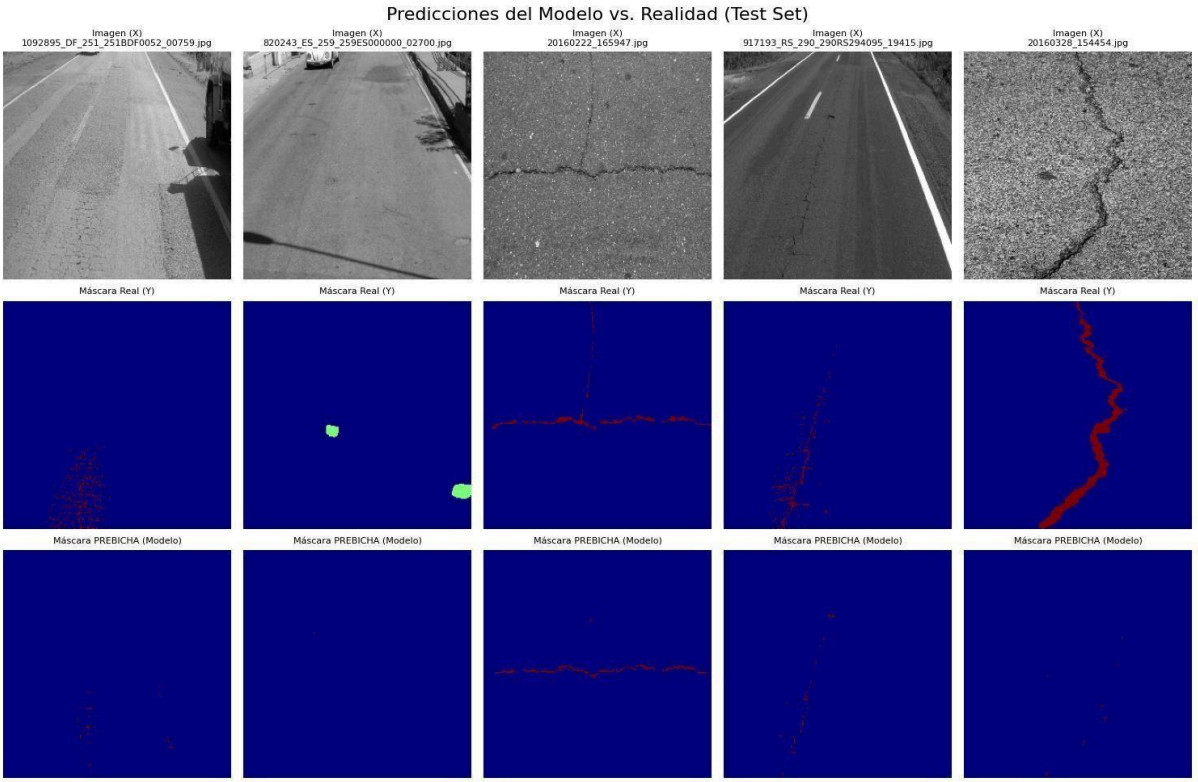


Figura 21. Resultados preliminares de la arquitectura U-net con 50 épocas y 167 iteraciones (Elaboración propia, 2025).

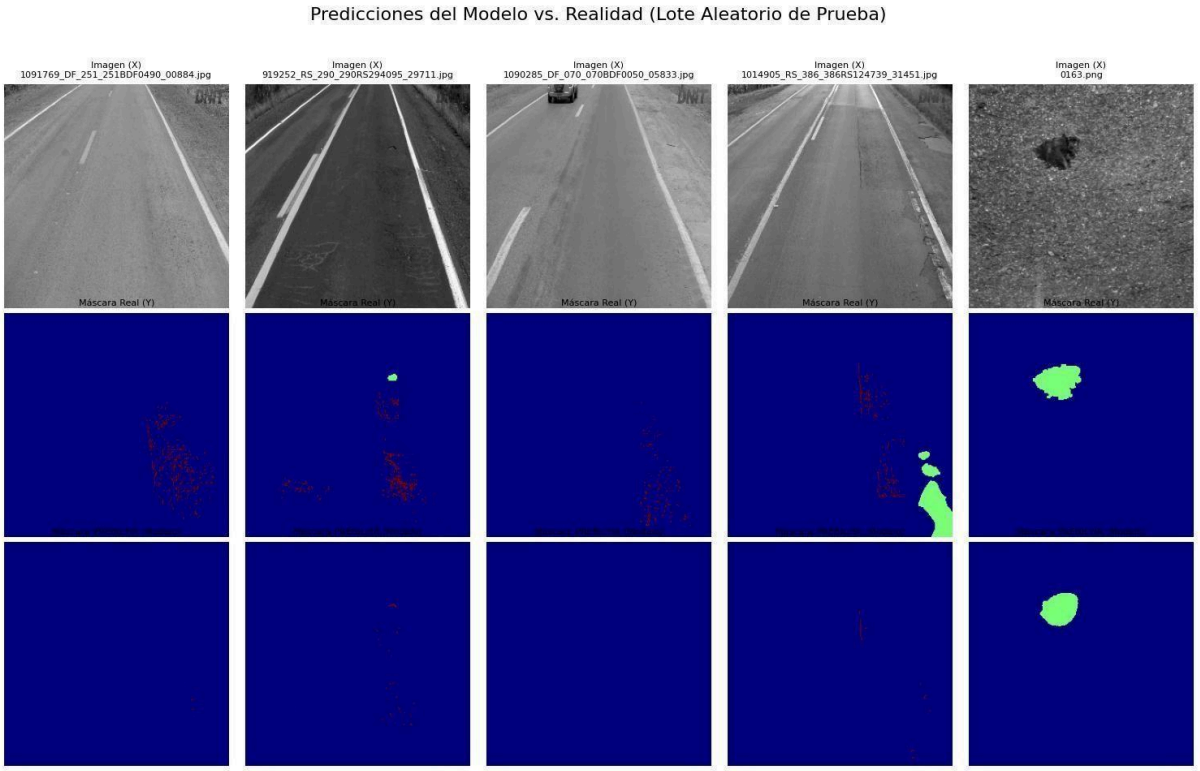


Figura 22. Resultados preliminares de la arquitectura U-net con 50 épocas y 167 iteraciones (Elaboración propia, 2025).

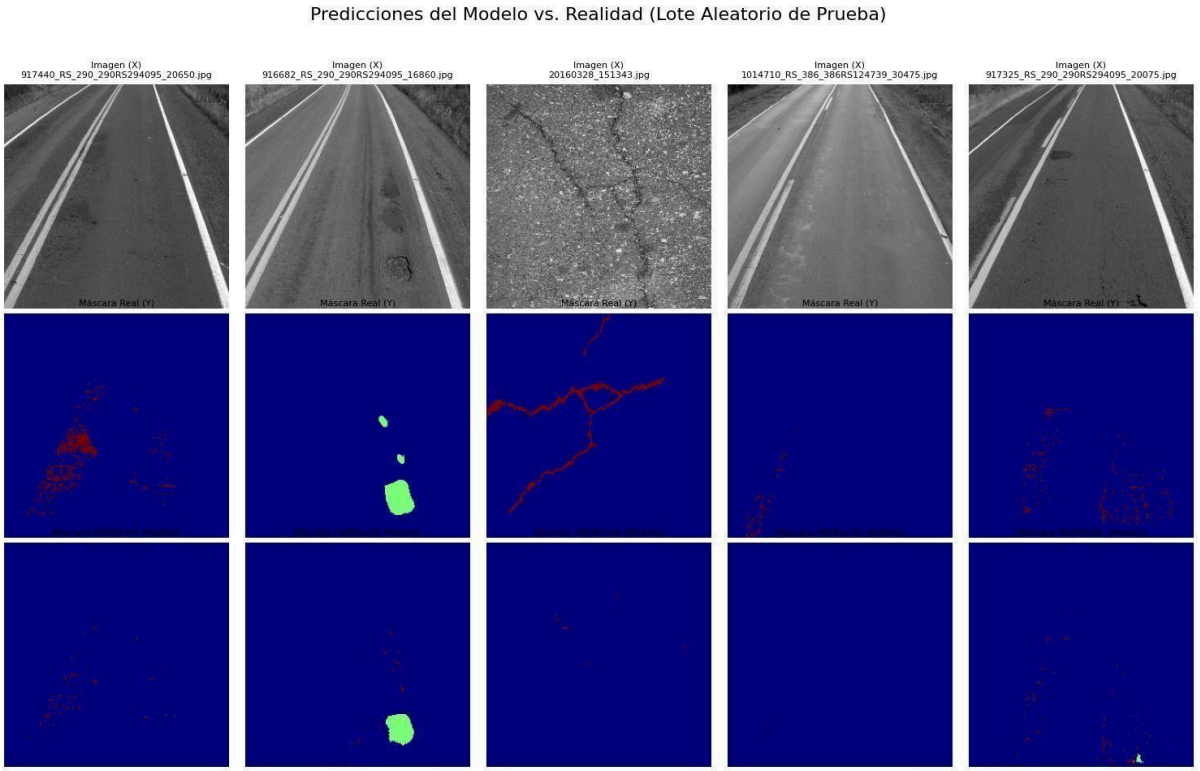


Figura 23. Resultados preliminares de la arquitectura U-net con 50 épocas y 167 iteraciones (Elaboración propia, 2025).

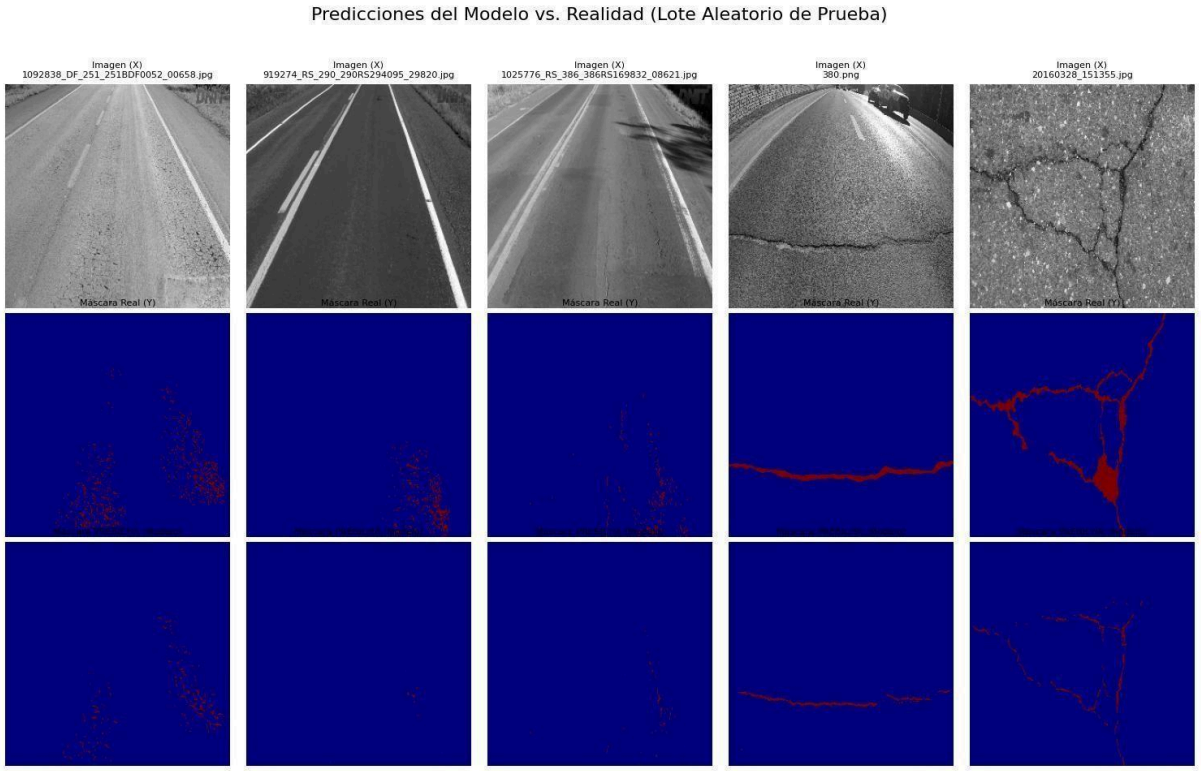


Figura 24. Resultados preliminares de la arquitectura U-net con 50 épocas y 167 iteraciones (Elaboración propia, 2025).

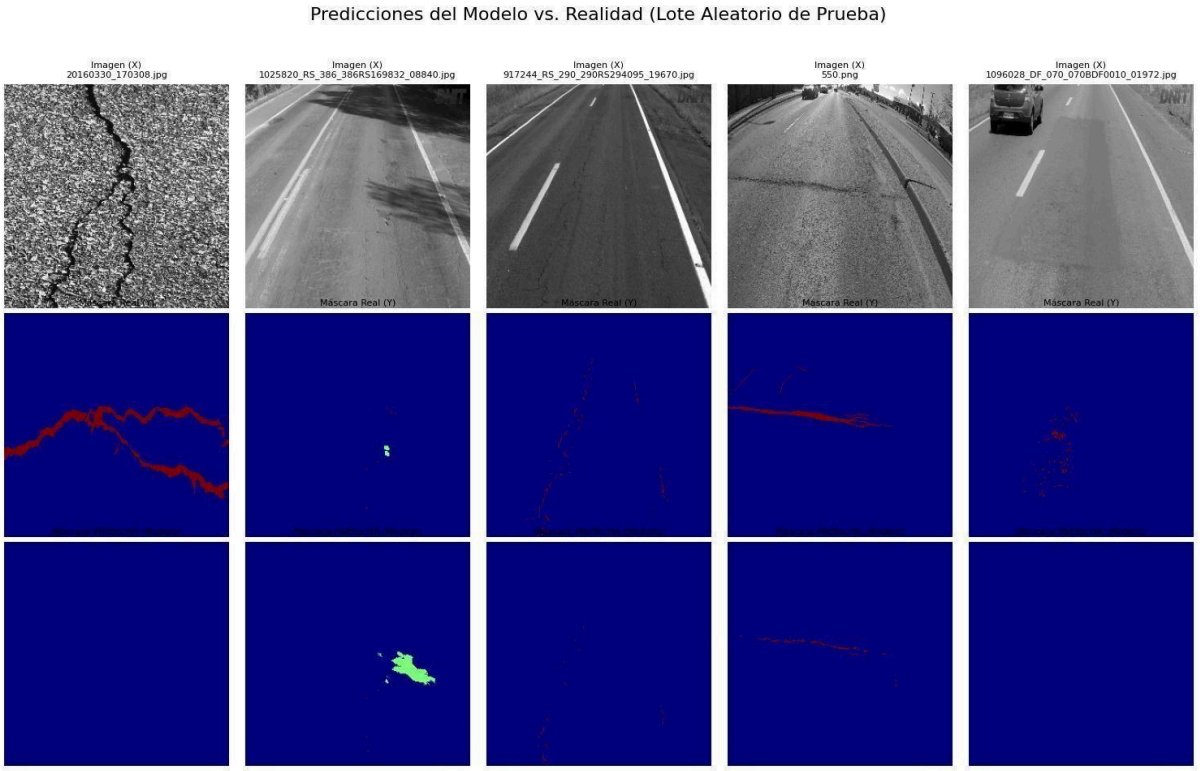


Figura 25. Resultados preliminares de la arquitectura U-net con 50 épocas y 167 iteraciones (Elaboración propia, 2025).

El análisis muestra tanto éxitos notables como fallas claras. Con esto se demostró una capacidad limitada para predecir precisamente, en pocos casos es igual a la máscara original, aunque logra diferenciar bien lo que es bache y fisura, también logra clasificar bien a las sombras como clase 0 (Aunque no perfectamente) y a los demás objetos que hay alrededor.

Es importante contar con estos resultados preliminares, ya que permiten identificar aspectos por mejorar. Aunque el proceso va bien encaminado, aún es necesario ajustar algunos hiperparámetros, optimizar el pipeline de preprocesamiento y considerar la posibilidad de incorporar nuevos datos con mayor contexto para evaluar el rendimiento del modelo. Estos aspectos se seguirán revisando y probando en etapas posteriores.

A partir de estos resultados preliminares, se evidenció que el modelo presentaba confusión entre clases, principalmente debido al desbalance extremo de píxeles (una gran proporción de

pavimento sano frente a una baja cantidad de fallas). Con base en esto, se plantearon varias estrategias que harán parte de la versión final del sistema.

En primer lugar, se propone la transición hacia un enfoque de tipo multi-experto. En lugar de utilizar una sola red para aprender todos los tipos de daño, se dividirá el problema en tareas más específicas. Para ello, se entrenarán dos modelos U-Net independientes y especializados: uno enfocado exclusivamente en la segmentación de fisuras (daños lineales finos) y otro orientado a la detección de baches (daños de tipo morfológico más severo). Posteriormente, las predicciones de ambos modelos se integrarán en un único diagnóstico, buscando un mejor desempeño conjunto.

Por otro lado, se identificó una limitación importante en la cantidad de datos disponibles, especialmente para la clase "bache", que cuenta con solo 710 imágenes. Para mitigar el riesgo de sobreajuste y evitar depender de bases de datos externas que no representen adecuadamente las condiciones del pavimento local, además de la limitada disponibilidad de datasets con imágenes y máscaras en formatos adecuados, se optó por implementar un modelo pre-entrenado (VGG16) como codificador para la detección de baches. Este enfoque se complementa con técnicas de aumento de datos (data augmentation) aplicadas en tiempo real, con el fin de incrementar la variabilidad del conjunto de entrenamiento y mejorar la capacidad de generalización del modelo.

H: Utilizar una función de pérdida basada en IoU permitirá mejorar el desempeño de la U-Net en la segmentación de fallas del pavimento, dado que esta métrica penaliza directamente los errores de superposición entre las regiones predichas y reales, alineando el proceso de entrenamiento con el criterio de evaluación final.	
H_0: No se observa mejora significativa al usar pérdida IoU.	**H_1:** El uso de pérdida IoU mejora el desempeño medido por IoU.

6. Resultados Finales

Se hizo una transición al Sistema Multi-Experto a partir de los resultados preliminares y el bajo rendimiento del modelo multiclase unificado, se determinó que la red sufría de confusión de clases. Para solucionar esto, se implementó un rediseño metodológico dividiendo el problema en un sistema "Multi-Experto": arquitecturas independientes especializadas en la morfología particular de cada daño (fisuras y baches) que posteriormente fusionan sus predicciones.

Para habilitar este enfoque, se desarrolló un script en Python encargado de procesar el dataset original, clasificando las imágenes a nivel de metadatos según el código de color de sus máscaras (rojo para fisuras, verde para baches, o ambos). Esta separación evidenció volúmenes de datos dispares, lo que obligó a tomar estrategias de optimización independientes para cada especialista.

Se usaron diferentes estrategias de optimización para cada especialista. Para el **especialista de fisuras,** la búsqueda de fuentes externas de datos permitió integrar un dataset de 7,732 imágenes y máscaras provenientes de *Hugging Face*, gestionado mediante un script de automatización hacia el entorno de Google Drive. Debido a que este nuevo conjunto de datos contenía grietas en diversos materiales (madera, paredes, concreto) y no exclusivamente en pavimento, se implementó la técnica de *Transfer Learning,* el cual es la reutilización del conocimiento visual de un modelo pre-entrenado (capaz de detectar formas y texturas complejas) para aplicarlo a un problema nuevo. En este contexto, equivale a tomar un sistema con un "ojo clínico" ya experto y calibrar específicamente para identificar patologías del pavimento, mitigando así la falta de bases de datos locales. Este con un enfoque en Adaptación de Dominio. Se inicializó la red con los pesos del modelo base pre-entrenado anterior y se ajustó (*fine-tuning*) con los datos locales, obligando al modelo a adaptar su capacidad general de detección de líneas delgadas específicamente a las texturas y condiciones del pavimento.

En cuanto al **especialista en baches**, la división del dataset local resultó en apenas 710 imágenes de baches. La búsqueda de datasets externos fue infructuosa debido a incompatibilidades de formato: la literatura actual tiende a usar formatos de coordenadas (*bounding boxes* tipo YOLO), los cuales encierran el daño en rectángulos, perdiendo la curvatura y el detalle a nivel de píxel que exige la arquitectura U-Net.

Para solventar esta carencia, se recurrió a la técnica de *Data Augmentation,* que consta de la expansión artificial de un conjunto de datos limitado mediante transformaciones geométricas aleatorias (rotaciones, espejos, zoom) sobre las imágenes originales. En la auscultación vial, esto simula observar la misma falla desde múltiples ángulos y condiciones de campo, obligando al modelo a aprender la morfología estructural real del daño. Este proceso se hizo sobre el conjunto de 568 imágenes de entrenamiento.

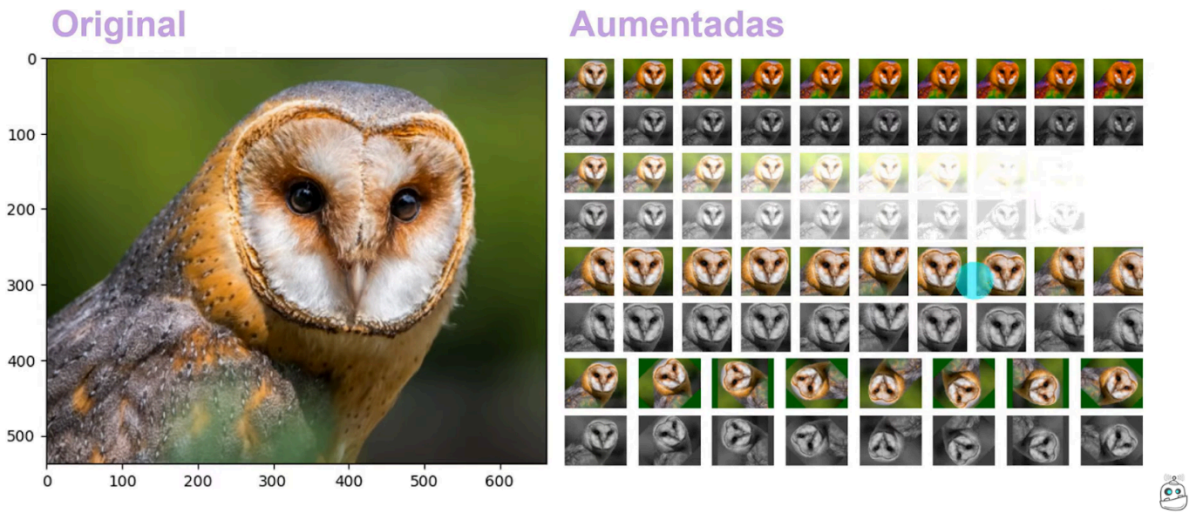


Figura 26. ***Ejemplo visual de transformaciones aplicadas mediante Data Augmentation(Código Máquina YouTube, 2024).***

Mediante variaciones geométricas aleatorias, el modelo generó un espacio muestral de 2,272 combinaciones únicas. A lo largo de las 15 épocas de entrenamiento, la red fue expuesta a

más de 8,500 iteraciones aumentadas, garantizando la generalización morfológica sin requerir datos sintéticos almacenados en disco.

El pipeline operó mediante dos conjuntos principales de transformaciones. El primero correspondió a las transformaciones geométricas, diseñadas para generar invarianza espacial frente a la orientación y la escala de los defectos viales. Durante cada época de entrenamiento, el algoritmo contó con una probabilidad del 50% de aplicar un reflejo horizontal y de manera independiente, una probabilidad del 50% de aplicar un reflejo vertical.

Adicionalmente, se configuró una rotación discreta aleatoria que modificó la orientación del tensor en múltiplos exactos de 90° (0°, 90°, 180° o 270°). Asimismo, con el propósito de simular variaciones en la distancia focal o en la altura de la cámara respecto al plano del pavimento, se incorporó una operación de acercamiento y recorte (Zoom / Crop).

Esta operación consistió en seleccionar una región aleatoria que conservaba entre el 80% y el 100% del área de la imagen original, la cual posteriormente era redimensionada mediante interpolación para ajustarse a las dimensiones estándar de la red neuronal (256 × 256 píxeles).

El segundo conjunto consistió en transformaciones fotométricas, las cuales se aplicaron exclusivamente a las imágenes de entrada y no a las máscaras binarias, con el propósito de emular las condiciones de iluminación incontroladas propias de la inspección en exteriores. Para ello, se introdujo una variación aleatoria en el nivel de brillo, ajustando los valores de los píxeles mediante un factor máximo de alteración de ±0.2. Esto permitió simular capturas realizadas en diferentes horas del día y bajo distintas condiciones lumínicas.

De forma complementaria, se aplicó un factor de multiplicación estocástico para modificar el contraste de la imagen dentro de un rango dinámico comprendido entre 0.8 (menor contraste) y 1.2 (mayor contraste). De esta manera, fue posible representar escenarios asociados tanto a condiciones de luz directa intensa como a zonas de sombra pronunciada proyectadas sobre la vía.

Por otro lado, con el fin de mitigar el sobreajuste y preparar al modelo para procesar imágenes de menor calidad o provenientes de sensores de bajo costo, se incorporó Ruido Gaussiano estocástico con una desviación estándar de 0.02 en el 50% de las iteraciones.

El rediseño arquitectónico fue acompañado de una optimización matemática. Al evaluar el rendimiento del modelo frente a la métrica estandarizada, **se rechaza la hipótesis nula H_0 y se acepta la hipótesis alternativa H_1**. La integración de la métrica IoU (combinada con el Coeficiente Dice) como función de pérdida dentro de la arquitectura U-Net forzó a la red a penalizar los errores de superposición. Los resultados evidencian una mejora sustancial frente a la función de entropía cruzada tradicional, evitando que el modelo colapsara hacia la clase mayoritaria (el pavimento sano).

6.1. Evaluación Cuantitativa y Comparativa

El rendimiento de cada modelo fue evaluado mediante cinco métricas estándar de segmentación semántica, calculadas a nivel de píxel sobre el conjunto de prueba:

Intersección sobre Unión (IoU): Mide la proporción entre el área de solapamiento y el área de unión entre la máscara predicha y la máscara real. Es la métrica más exigente porque penaliza simultáneamente los falsos positivos y los falsos negativos.

$$IoU = \frac{TP}{TP+FP+FN}$$

Coeficiente Dice: Equivalente al F1 calculado a nivel de píxel. Es más sensible que el IoU ante clases pequeñas porque pondera doblemente los verdaderos positivos.

$$\text{Dice} = \frac{2 * TP}{2 * TP + FP + FN}$$

Precisión: Proporción de píxeles clasificados como falla que efectivamente corresponden a una falla real. Mide la tasa de falsos positivos del modelo.

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

Recall: Proporción de píxeles de falla reales que el modelo detectó correctamente. Mide la tasa de falsos negativos, que en el contexto vial es la métrica de mayor impacto operativo.

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

F1-Score: Media armónica entre Precisión y Recall, utilizado como métrica de síntesis cuando existe un compromiso entre ambas.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recal}$$

La superioridad del enfoque propuesto se valida al contrastar las métricas globales obtenidas sobre el conjunto de pruebas (Test). A continuación, se detalla el salto de rendimiento entre el intento inicial unificado y los especialistas finales, aunque cabe resaltar que en este apartado no se tomó el F1 Score:

RESULTADOS U-NET 30 EPOCAS **BACHES Y GRIETAS**				
	Métrica IoU	Métrica Dice	Precisión	Recall
Fondo	0.9860	0.9930	0.9883	0.9977
Bache	0.5564	0.7150	0.7904	0.6527

Agrietamiento	0.1300	0.2301	0.6713	0.1388

Tabla 2. Resultados del modelo U-Net entrenado durante 30 épocas para la segmentación de baches y grietas en pavimentos (Elaboración propia, 2026).

RESULTADO MODELO ESPECIALISTA EN FISURAS **AGRIETAMIENTO** 15 EPOCAS				
	Métrica IoU	Métrica Dice	Precisión	Recall
Fondo	0.9758	0.9878	0.9866	0.9889
Agrietamiento	0.1974	0.3296	0.3512	0.3106

Tabla 3. Resultados del modelo U-Net especialista en agrietamiento entrenado durante 15 épocas para la segmentación de Agrietamientos en pavimentos (Elaboración propia, 2026).

RESULTADO MODELO ESPECIALISTA EN FISURAS **BACHES** 15 EPOCAS				
	Métrica IoU	Métrica Dice	Precisión	Recall
Fondo	0.9878	0.9938	0.9908	0.9969
Baches	0.6124	0.7596	0.8613	0.6795

Tabla 4. Resultados del modelo U-Net especialista en baches entrenado durante 15 épocas para la segmentación de baches en pavimentos (Elaboración propia, 2026).

Como se observa en las tablas comparativas, la desacoplación de las arquitecturas y la implementación de la nueva función de pérdida incrementaron drásticamente tanto la precisión

como el nivel de intersección (IoU), validando la viabilidad del sistema Multi-Experto para la segmentación de fallas viales complejas.

Para complementar el rendimiento estadístico, se evaluó cualitativamente el modelo superponiendo las predicciones (máscaras generadas) sobre fotografías reales del pavimento en escenarios de prueba no vistos durante el entrenamiento. Como se evidenció en los resultados preliminares, el modelo unificado fallaba en distinguir los límites del daño. Por el contrario, la arquitectura Multi-Experto demostró una notable precisión geométrica.



Figura 27. ***Resultado Modelo U-net especialista en Agrietamiento (Elaboración propia, 2026).***



Figura 28. ***Resultado Modelo U-net especialista en Agrietamiento (Elaboración propia, 2026).***



Figura 29. ***Resultado Modelo U-net especialista en baches (Elaboración propia, 2026).***



Figura 30. ***Resultado Modelo U-net especialista en baches probada con imagen de Google (Elaboración propia, 2026).***

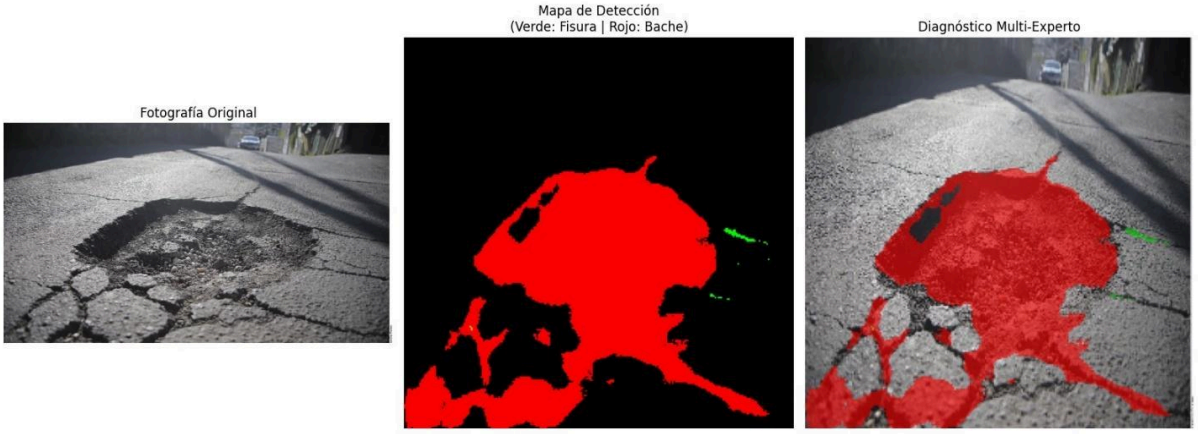


Figura 31. ***Resultado Modelo U-net especialista en baches y agrietamiento juntas con imágenes de Google (Elaboración propia, 2026).***

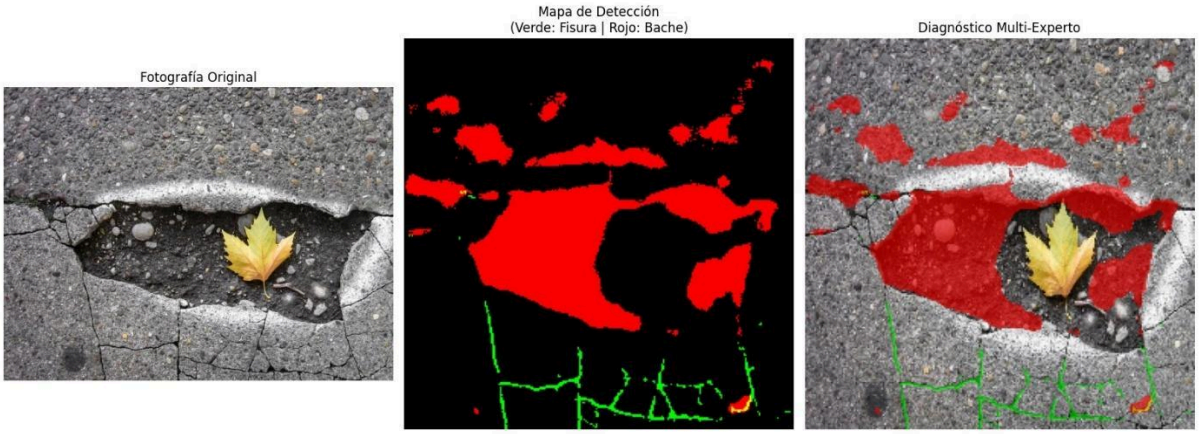


Figura 32. ***Resultado Modelo U-net especialista en baches y agrietamiento juntas con imágenes de Google (Elaboración propia, 2026).***

Cabe destacar la robustez del especialista en baches (VGG 16) frente a fenómenos de oclusión. Durante las pruebas visuales, el modelo fue capaz de identificar los bordes asfálticos de un bache severo ignorando objetos extraños en su interior (como hojas o basura en este ejemplo), lo cual demuestra que la red neuronal extrajo exitosamente las características morfológicas profundas del daño y no dependió únicamente de variaciones simples de color o iluminación.

Asimismo, el especialista en fisuras logró aislar las ramificaciones más delgadas del pavimento, cumpliendo con el objetivo de un diagnóstico temprano para mantenimiento preventivo.

6.2. Resultados Finales

A lo anteriormente mencionado se le hicieron ajustes y mejoras para poder presentar los mejores resultados posibles, así que a continuación, se presentarán los ajustes de los modelos finales:

En la parte de fisuras se agregó un mecanismo de atención o mejor conocido como attention gate, que fue el encargado hacer que el modelo ignorara el fondo que generaba ruido en la red y así pudiera enfocarse más en detallar mejor las delgadas líneas que se presentaban en las grietas. Hubo un gran reto en el **desbalance de clases**, ya que normalmente el pavimento sano representa aproximadamente el 95% de cada imagen mientras que el 5% restante de la imagen es la falla. Para esto se optó por implementar **Focal Loss y Tversky Loss.**

- **Focal Loss:** Era el encargado de hacer que el modelo se enfocara en la clase minoritaria, en los píxeles más complejos de la imagen, los cuales eran las fallas que hubiese.

- **Tversky Loss:** Con su parámetro de **α= 0.7**, lo que hace es penalizar matemáticamente de forma agresiva si el modelo produce un falso negativo o sea si ignora una grieta. Se usó un **α** = 0.7 porque viendo la fórmula matemática del Tversky Loss:

$$T(\alpha, \beta) = \frac{TP}{TP + \alpha \cdot FN + \beta \cdot FP}$$

α: Es el peso del castigo por los Falsos Negativos (grietas reales que el modelo no vio).

β: Es el peso del castigo por los Falsos Positivos (pavimento sano o manchas que el modelo confundió con grietas).

Al aumentar el valor de $\alpha = 0.7$, se decidió penalizar con mayor peso los falsos negativos, basándose en el análisis de dos situaciones reales pero hipotéticas (para mayor entendimiento) utilizadas como referencia:

Caso A (Falso Positivo): El modelo detecta una grieta donde en realidad hay una sombra, una mancha de aceite o pavimento húmedo. **Consecuencia:** Se envía un operario a inspeccionar, quien confirma que no existe daño y continúa su labor. Esto implica una pérdida de tiempo operativa, pero con impacto limitado y bajo riesgo a futuro.

Caso B (Falso Negativo): Existe una grieta capilar en formación, pero el modelo la clasifica como pavimento en buen estado. **Consecuencia:** La grieta no es atendida, permite la infiltración de agua y acelera el deterioro estructural, pudiendo evolucionar en poco tiempo a un bache significativo que genera costos elevados de reparación e incluso riesgos de accidente.

A partir de este análisis, se prioriza la penalización del Caso B, dado que sus consecuencias son considerablemente más críticas en el contexto del mantenimiento vial. Fue así que la elección de $\alpha=0.7$ no es arbitraria, sino que responde a un criterio de riesgo propio del dominio del problema. En aplicaciones de inspección de pavimentos, el costo asociado a un falso negativo (no detectar una fisura temprana) es estructural y económicamente superior al de un falso positivo (realizar una inspección innecesaria).

Por ello, al emplear un Índice de Tversky asimétrico con $\alpha=0.7$ y $\beta=0.3$, la función de pérdida penaliza de forma más severa la omisión de grietas, favoreciendo un incremento en el recall (sensibilidad), sin deteriorar significativamente la precisión del modelo.

También, se implementó una función encargada de ayudar en la robustez de los datos con **augmentation sincronizado.** Esta función aplico 8 transformaciones matemáticas en tiempo real (zoom, recortes, flips, ruido gaussiano, cambios de brillo) garantizando que la máscara (Y) sufra exactamente la misma deformación que la foto (X). Esto incrementó significativamente el volumen del conjunto de datos original, mitigando efectivamente el riesgo de sobreajuste. Así se logró tener una mayor cantidad de datos y en cuanto a los anteriores resultados, esta vez se verificó que las transformaciones se hicieron en simultáneo con las máscaras.

Asimismo, se implementó algo muy interesante, la **búsqueda de umbral óptimo**, que se encargará de evaluar automáticamente cual es el mejor punto de corte basándose en el set de validación para maximizar el IoU. Además, se usó **TTA (Test-Time Augmentation),** cuando ingrese una imagen nueva, el sistema final no se conforma con verla una vez. La rota y espejea 6 veces, hace 6 predicciones distintas y entrega el promedio. Esto reduce los puntos ciegos del modelo.

Aparte, se implementó en el código la visualización de la máscara que se generará tanto en las fisuras como en los baches de la imagen que entra en el modelo. La salida original del modelo es una matriz binaria tras aplicar el umbral de decisión. Para la visualización de los resultados, se aplica un mapa de pseudocolor (usando colormap) sobre los píxeles clasificados como positivos. Se seleccionó una paleta de colores cálidos (tonos amarillos y naranjas) por convención de diseño de interfaces, ya que estos colores generan un alto contraste con la escala de grises del pavimento y dirigen inmediatamente la atención hacia las zonas de daño crítico, facilitando la interpretación operativa de la imagen. En terminos mas breves, Lo amarillo que se resalata en las imagenes es estrictamente la zona donde la red neuronal superó el umbral operativo definido para cada experto, confirmando con alta seguridad: "aquí hay daño".

También, se calcularon unos porcentajes de área afectada en cada imagen por cada especialista (Bache y fisuras). El porcentaje de área afectada se calcula mediante una proporción directa de píxeles. Dado que la red estandariza todas las imágenes de entrada a una resolución de 256x256, cada imagen contiene un total exacto de 65,536 píxeles. El porcentaje se obtiene sumando la cantidad total de píxeles que la red clasificó positivamente como falla (fisura o bache) y dividiendo este valor sobre el total de píxeles de la imagen. Esta métrica representa la densidad del daño visible en el encuadre capturado y se representa con la siguiente fórmula:

$$Porcentaje_Afectacion = \left(\frac{\sum Píxeles_Falla}{Ancho \times Alto} \right) \times 100$$

De igual forma, se definieron unas clasificaciones de severidad para mostrar la gravedad de cada falla en el pavimento. Cabe resaltar que la clasificación de severidad no es una predicción directa de la red neuronal, sino una categorización determinista basada en el porcentaje de área afectada (densidad del daño). Para traducir los resultados matemáticos de la segmentación a un formato de diagnóstico vial procesable, se definieron los siguientes umbrales operativos:

- **Sin falla:** Afectación menor al 0.5% (margen de tolerancia para ruido o falsos positivos aislados).
- **Ligera:** Afectación entre 0.5% y 5.0% del área (daños superficiales o incipientes).
- **Moderada:** Afectación entre 5.0% y 15.0% (daños en desarrollo que requieren monitoreo).
- **Severa:** Afectación superior al 15.0% (daño estructural evidente que requiere mantenimiento correctivo prioritario).

Todo esto junto dio como resultados finales las siguientes pruebas que se hicieron con imágenes tomadas por el autor y algunas de internet. Cabe resaltar que estas explicaciones y muestras de los resultados fueron únicamente con 3 imágenes por categoría para la facilidad visual del lector:

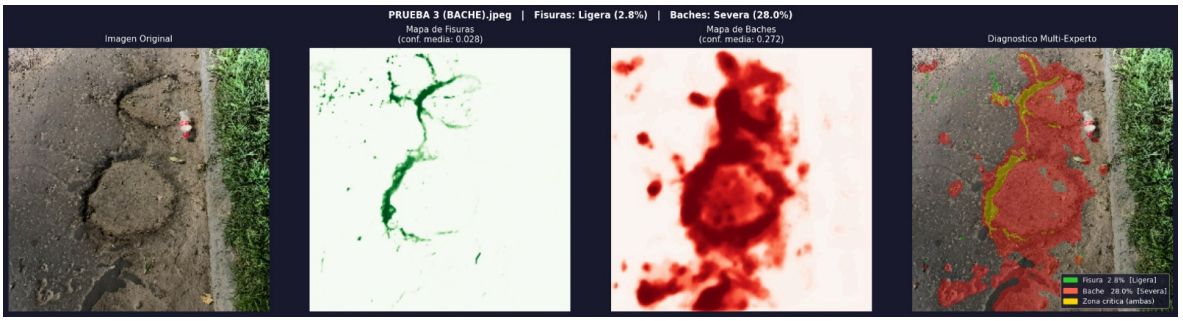
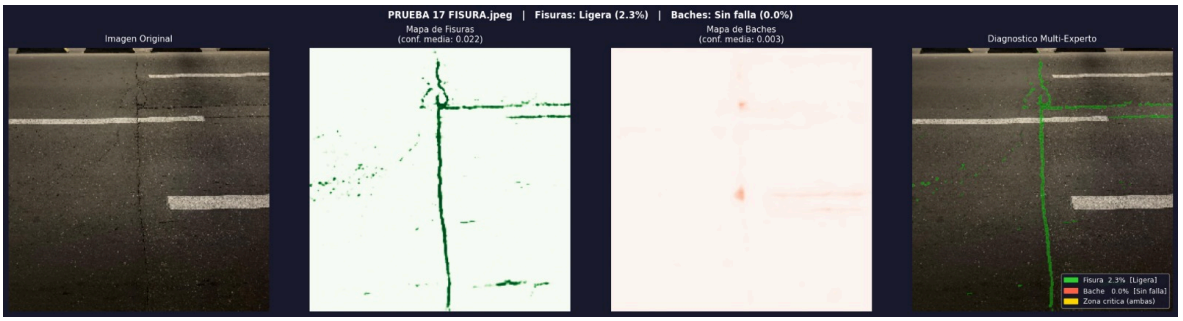
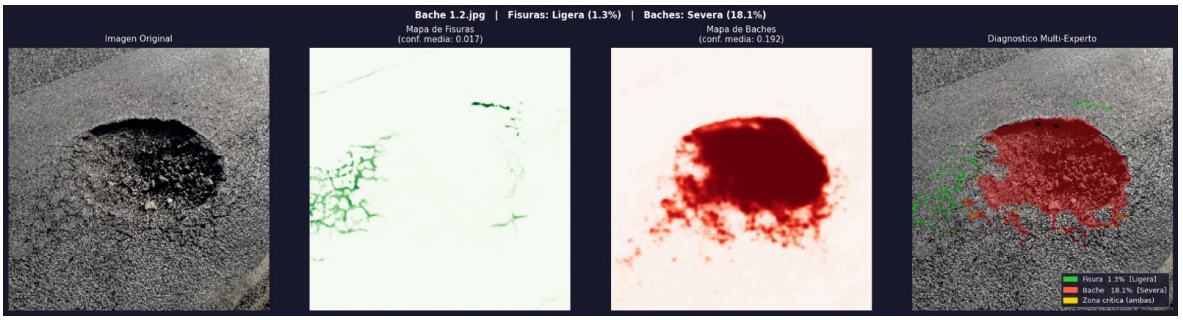
RESULTADOS DE ÉXITO
 Fotografía tomada por el autor
 Fotografía tomada por el autor
 Imagen de prueba tomada de Letras Libres, 2022.

Tabla 5. Evidencias de los resultados. (Elaboración propia, 2026).

En la categoría de resultados de éxito se pusieron algunos ejemplos donde el modelo se desempeñó de la mejor forma, mostró que es capaz de reconocer los dos tipos de fallas con

imágenes capturadas en el mundo real, sin confusiones. En la primera imagen se observa una segmentación precisa de la falla predominante (el bache en ese caso). El modelo especialista correspondiente logra delimitar la morfología del daño aislando el ruido del fondo, como las variaciones de iluminación, sombras y la textura rugosa del pavimento sano. Es notable que, ante este daño específico, el otro modelo de la arquitectura multi-experto se mantiene parcialmente inactivo. Esto demuestra una excelente capacidad del sistema para evitar falsos positivos cruzados; es decir, la red no confunde la profundidad y densidad de un bache con el patrón ramificado y lineal de una grieta. Igualmente toca resaltar que se equivoca un poco con un pedazo de la pintura amarilla que hay y la clasifica como bache.

En la segunda imagen, se evidencia la efectividad de la arquitectura ante un escenario de daño más pronunciado y delgado. La máscara binaria generada cubre la zona afectada de forma sólida y contigua, lo cual refleja que el extractor de características y los mecanismos de atención calcularon una confianza probabilística alta y sostenida (muy por encima del umbral operativo de 0.5) en toda el área crítica, sin fragmentar la predicción.

Finalmente, en la tercera imagen, el sistema valida por completo el enfoque multiexperto propuesto en esta investigación al procesar un entorno visual complejo. Los resultados muestran cómo los dos modelos especialistas logran operar en paralelo de forma exitosa sin solaparse erróneamente: la red aísla la depresión principal asignándole a la máscara de baches, al mismo tiempo que el segundo modelo identifica y mapea las fisuras adyacentes. Este nivel de detalle confirma que la arquitectura logra generalizar adecuadamente los bordes difusos y comprende las fronteras físicas del pavimento, distinguiendo exactamente dónde termina un tipo de daño estructural y comienza el otro.

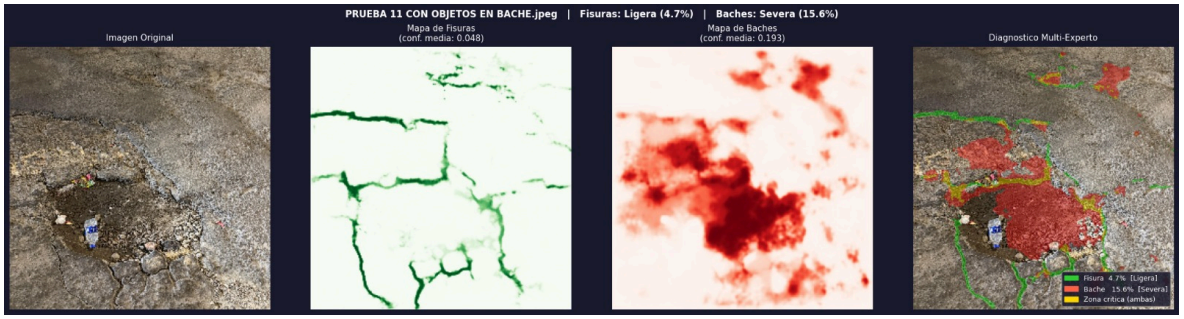
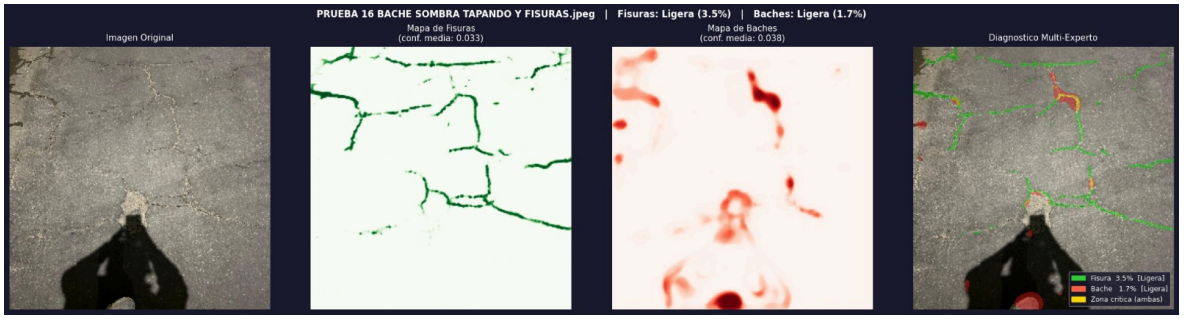
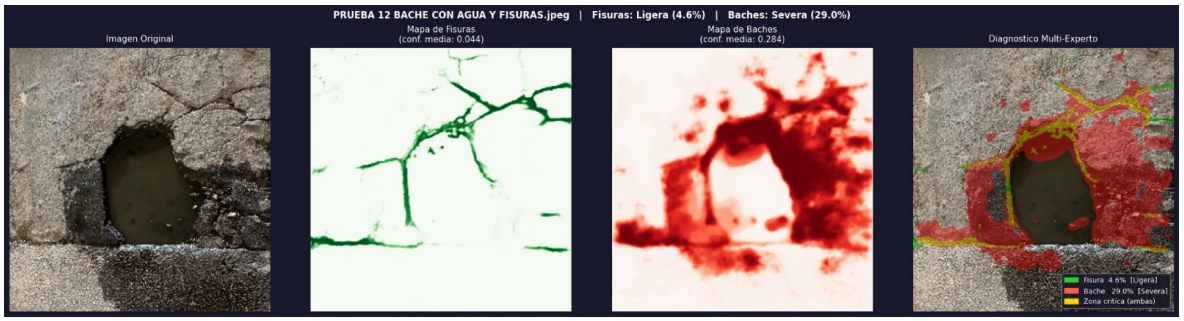
RESULTADOS Y LIMITACIONES DEL MODELO
 Fotografía tomada por el autor
 Fotografía tomada por el autor
 Fotografía tomada por el autor

Tabla 6. Evidencias de los resultados. (Elaboración propia, 2026).

El análisis de casos complejos revela ciertas limitaciones de la arquitectura frente a factores ambientales y oclusiones. Por un lado, la presencia de agua, escombros o elementos ajenos dentro de un bache altera la continuidad de los gradientes de textura que la red utiliza para la extracción de características. Esto provoca una caída en la confianza probabilística del modelo, generando segmentaciones parciales o falsos negativos, al omitir el área total del daño estructural. Por otro lado, el sistema demostró susceptibilidad a falsos positivos derivados de variaciones abruptas en el contraste. Elementos como sombras proyectadas por el propio autor en este caso, o incluso objetos intrusivos en el encuadre periférico (como el calzado del operario al capturar la

imagen), fueron clasificados erróneamente como anomalías debido a la similitud de su contorno oscuro con la morfología de un bache. El mapeo de estas vulnerabilidades visuales resulta fundamental para el análisis de error, estableciendo una hoja de ruta clara para trabajos futuros, orientada a la inclusión sistemática de estos ruidos ambientales en las estrategias de Data Augmentation para robustecer la generalización del modelo.

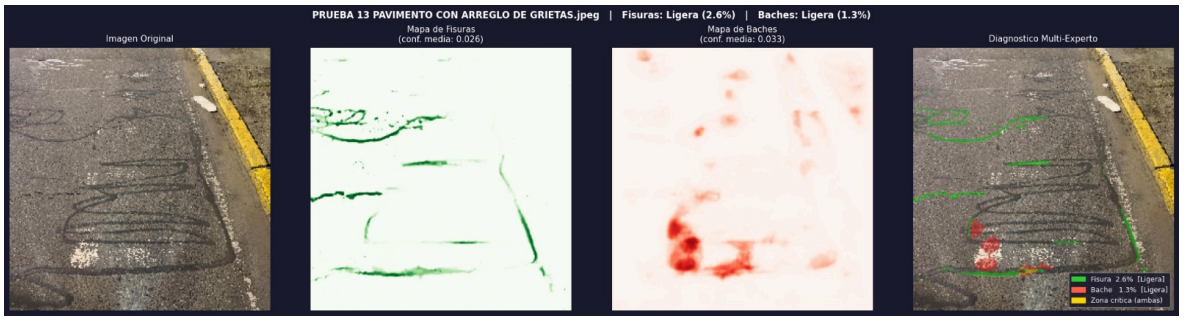
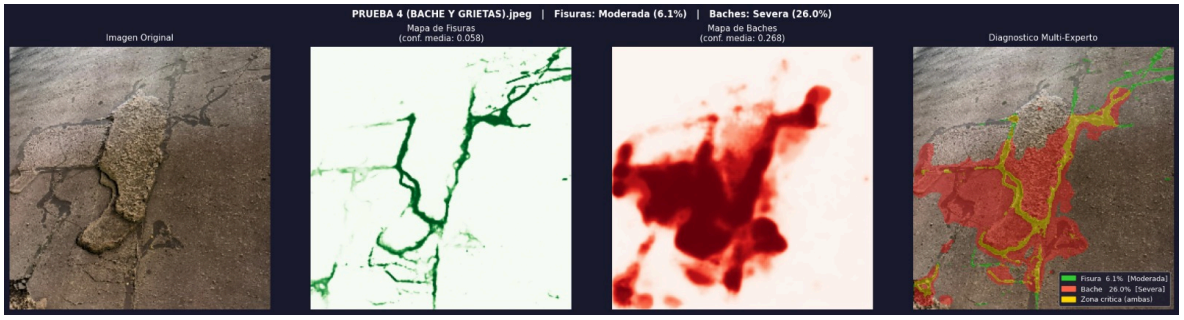
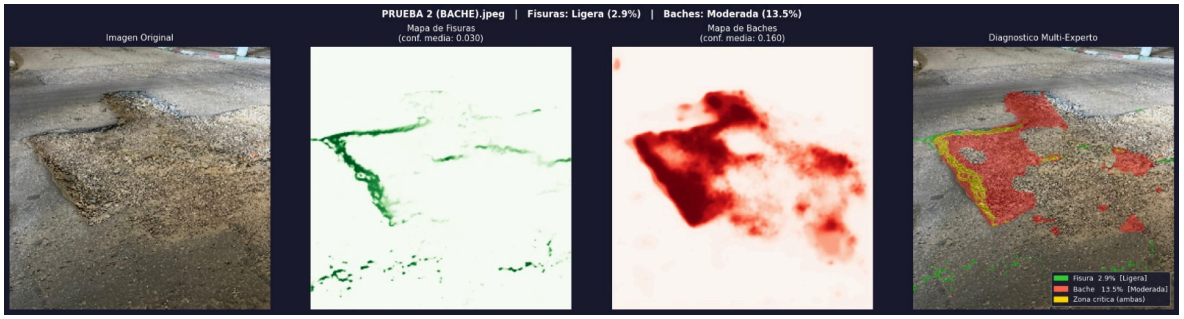
RESULTADOS Y ERRORES DEL MODELO
 Fotografía tomada por el autor
 Fotografía tomada por el autor
 Fotografía tomada por el autor

Tabla 7. Evidencias de los resultados. (Elaboración propia, 2026).

En la primera imagen se observa un caso claro de falso positivo, donde el modelo identifica fisuras en un pavimento que en realidad corresponde a una intervención previa de sellado. Este comportamiento evidencia que el modelo está aprendiendo patrones visuales basados

principalmente en la forma y la textura, especialmente líneas alargadas que se asemejan a grietas, pero no logra distinguir entre una fisura activa y una reparación. Desde un punto de vista técnico, esto refleja una limitación en la capacidad de generalización semántica, ya que ambas situaciones comparten características visuales como contraste y bordes definidos. El bajo porcentaje de detección también sugiere que el modelo tiene incertidumbre en este tipo de escenarios, lo que confirma que se trata de una clasificación ambigua.

En la segunda imagen el problema se intensifica, ya que el modelo no solo confunde reparaciones con fisuras, sino que además clasifica como bache una zona que corresponde a pavimento deteriorado pero no necesariamente a un bache estructural completo. Esto indica una tendencia a sobre-detectar daños cuando existen patrones de alta irregularidad visual. En este caso, el modelo parece priorizar características locales como la rugosidad o la discontinuidad de la superficie, sin considerar adecuadamente el contexto global de la imagen. Como resultado, asigna una severidad alta que puede no corresponder con la realidad del daño. Este comportamiento sugiere un sesgo hacia texturas complejas, lo cual es común en modelos entrenados con datos donde estas características están fuertemente asociadas a la clase de bache.

Mientras que, en la tercera imagen el modelo muestra un mejor desempeño en la identificación del bache, logrando ubicar correctamente la zona afectada. Sin embargo, se evidencia un problema de subsegmentación, ya que solo detecta parcialmente el área del daño y no logra cubrir completamente sus límites. Este tipo de error suele estar relacionado con umbrales de activación elevados o con limitaciones en la consistencia espacial de la segmentación. Aunque el modelo reconoce correctamente la presencia del bache, la delimitación incompleta impacta directamente en la estimación del área afectada y, por ende, en la clasificación de severidad.

Asimismo, con el fin de proporcionar una referencia visual más clara, en la Figura 33 se presentan de manera independiente y comparativa las topologías de las dos redes U-Net que

conforman el sistema Multi-Experto: el especialista en baches y el especialista en fisuras. Esta representación permite identificar y diferenciar las etapas de contracción (Encoder), el cuello de botella y la expansión (Decoder), así como las modificaciones específicas implementadas en las conexiones residuales, los pipelines de entrada y las funciones de costo empleadas en cada modelo.

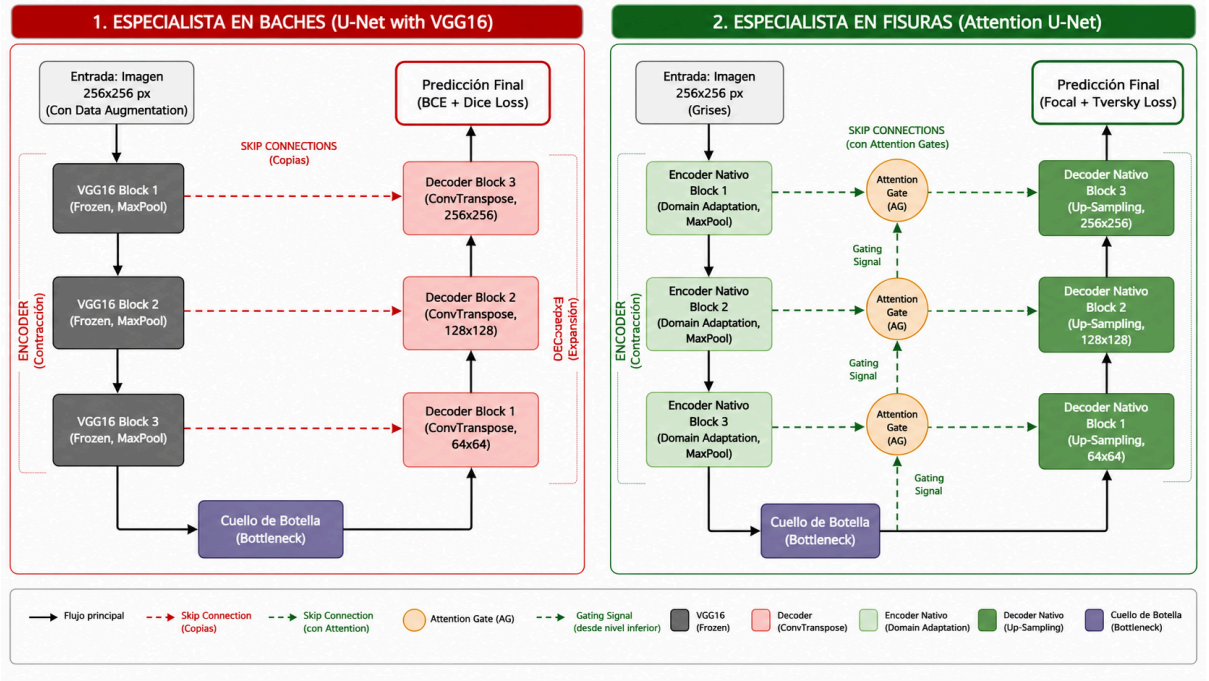


Figura 33.Representación gráfica de las arquitecturas(Experto en fisuras y experto en baches). Elaboración Propia

Validación Cruzada

Con el fin de obtener una estimación robusta del rendimiento generalizable de cada especialista, se aplicó una validación cruzada de k=5 folds sobre el conjunto de datos completo de cada clase. Este procedimiento divide aleatoriamente el dataset en cinco particiones mutuamente excluyentes, entrenando el modelo en cuatro de ellas y evaluándolo en la restante, repitiendo el proceso cinco veces de modo que cada imagen actúa exactamente una vez como dato de prueba. Los resultados reportados corresponden a la media y desviación estándar de las métricas de segmentación a lo largo de los cinco folds.

El modelo de fisuras fue evaluado sobre 2.852 imágenes, con particiones de aproximadamente 2.281 imágenes para entrenamiento y 571 para prueba en cada fold. Los resultados obtenidos fueron:

Métrica	Media	± Std
IoU	0.1862	± 0.0071
Dice	0.2955	± 0.0098
Precisión	0.2955	± 0.0230
Recall	0.4615	± 0.0523
F1	0.3574	± 0.0108

Tabla 8. Tabla de resultados métricas de evaluación. (Elaboración propia, 2026).

El indicador más relevante para interpretar estos resultados es la bajísima desviación estándar en todas las métricas: la variación del IoU entre folds es de apenas ±0.0071, lo que demuestra que el modelo aprende patrones consistentes independientemente de qué fracción del dataset se use para entrenamiento. Esto se aprecia visualmente en la gráfica (parte inferior) de IoU por fold, donde la curva azul permanece prácticamente horizontal entre 0.176 y 0.196 a lo largo de los cinco folds.

En términos absolutos, el IoU de 0.186 refleja la dificultad intrínseca de la detección de fisuras: se trata de estructuras lineales delgadas que ocupan una fracción muy pequeña del área total de la imagen, generando un desbalance extremo entre píxeles de falla y de pavimento sano. Esta condición penaliza directamente el IoU, que es sensible a la proporción de píxeles correctamente segmentados sobre el total de la unión. El valor de Recall de 0.461 indica que el modelo logra detectar casi la mitad de los píxeles de fisura presentes, lo cual es un resultado razonable dado el desbalance, y es consistente con la estrategia de la función de pérdida combinada (Focal Loss + Tversky Loss con α=0.7) diseñada precisamente para priorizar la sensibilidad sobre la precisión.

De otro modo, el modelo de baches fue evaluado sobre 710 imágenes, con particiones de 568 imágenes para entrenamiento y 142 para prueba. Los resultados obtenidos fueron:

Métrica	Media	± Std
IoU	0.4032	± 0.0363
Dice	0.5155	± 0.0432
Precisión	0.6982	± 0.0875
Recall	0.5559	± 0.0676
F1	0.6111	± 0.0349

Tabla 9. Tabla de resultados métricas de evaluación. (Elaboración propia, 2026).

El especialista en baches alcanza un IoU medio de 0.403, más del doble que el especialista en fisuras, lo que refleja la naturaleza morfológicamente distinta de esta clase: los baches son regiones compactas de mayor área superficial, lo que reduce el desbalance de clases y facilita la segmentación. La Precisión de 0.698 indica que cuando el modelo predice un bache, acierta en aproximadamente el 70% de los píxeles, evidenciando que el encoder VGG16 pre-entrenado en ImageNet aporta características de alto nivel que reducen las falsas detecciones.

La desviación estándar es más alta que en fisuras, especialmente en Precisión (±0.0875), lo que puede atribuirse al tamaño reducido del dataset de baches (710 imágenes frente a 2.852 de fisuras): con particiones de solo 142 imágenes por fold de prueba, la varianza estadística es naturalmente mayor. El boxplot (que se encuentra en la parte inferior) confirma esta observación, mostrando una distribución más dispersa para los baches que para las fisuras, e incluso dos valores atípicos en las métricas de Precisión y Recall.

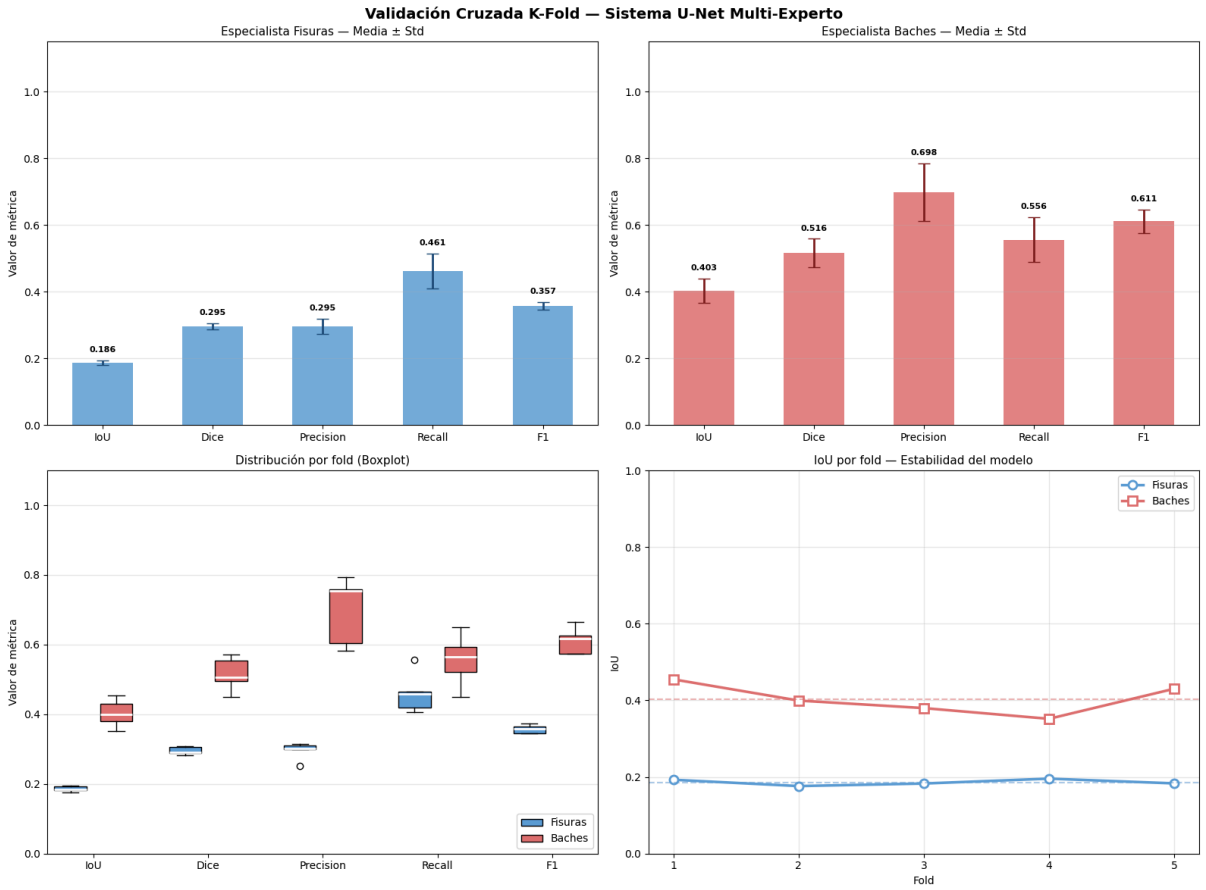


Figura 34. Gráficas sobre resultados que se obtuvieron en validación cruzada. Elaboración Propia

La comparación entre ambos especialistas evidencia dos retos muy diferentes con resultados bastante desiguales. Las fisuras presentan alta estabilidad (std baja) pero rendimiento absoluto limitado por el desbalance de clases y la delgadez de las estructuras a segmentar. Mientras que los baches presentan rendimiento absoluto superior pero mayor varianza, atribuible al menor volumen de datos disponibles.

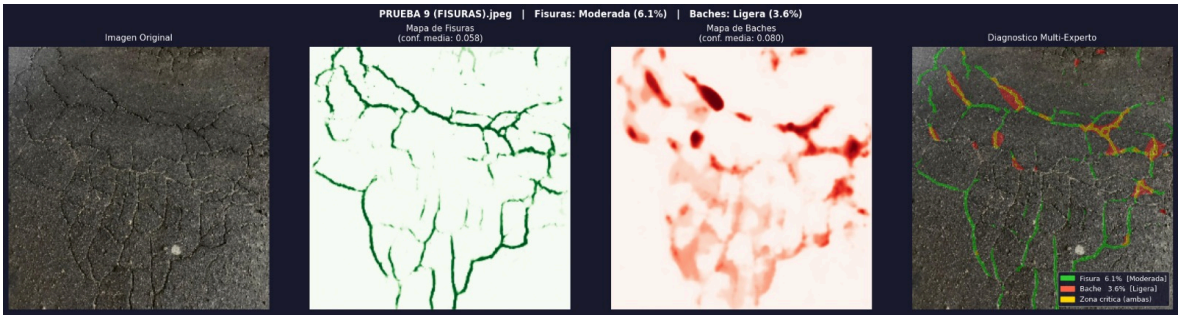
Ambos resultados son coherentes con lo reportado en la literatura de segmentación vial: modelos de segmentación semántica sobre fisuras en pavimento típicamente reportan IoU entre 0.15 y 0.35 en condiciones de datos limitados y alto desbalance, mientras que la detección de baches suele alcanzar rangos de IoU entre 0.35 y 0.65 dependiendo del dataset. Los valores obtenidos en este trabajo se sitúan dentro de dichos rangos, lo que valida la solidez metodológica

del sistema propuesto.

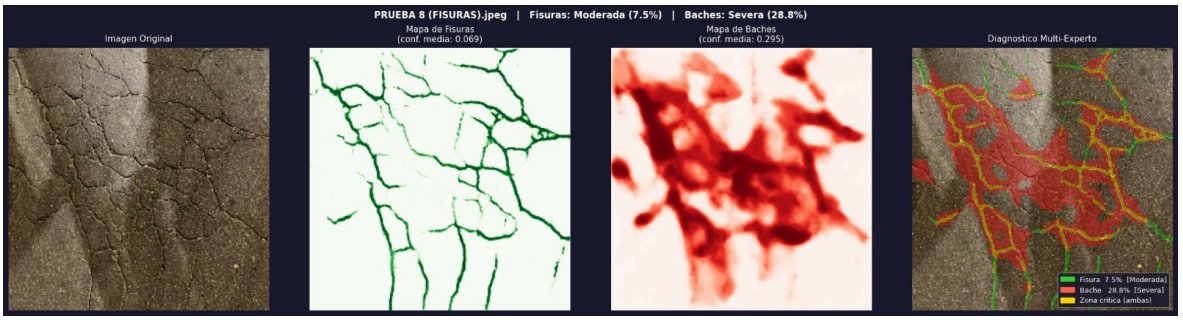
Finalmente, la estabilidad del IoU de fisuras a lo largo de los cinco folds (rango de solo 0.019 entre el valor mínimo y máximo) confirma que el modelo no sobre ajusta a ninguna partición particular del dataset, y que la estrategia de transfer learning desde el dataset HuggingFace seguida de fine-tuning local produce un aprendizaje generalizable. Reforzando y pudiendo afirmar que el overfitting es nulo, tanto por los resultados que muestra la validación cruzada como los resultados visuales con imágenes externas a los dataset.

Por último se dejarán otros resultados visuales que se tuvieron con las imágenes tomadas manualmente:

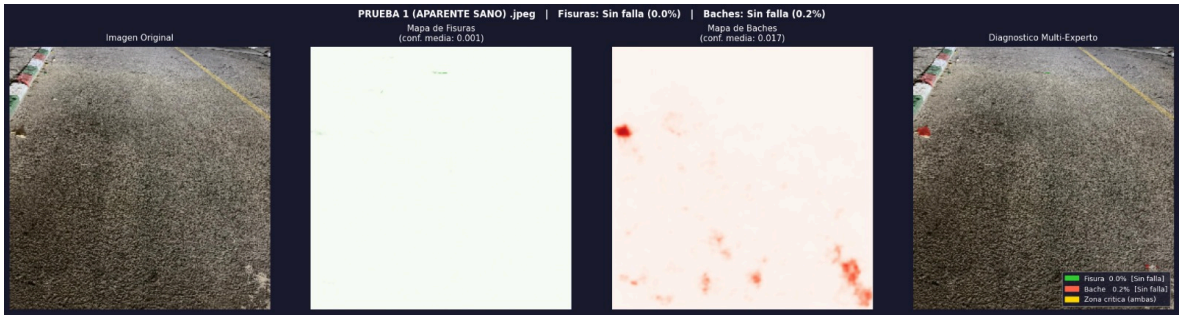
EXTRA: RESULTADOS VISUALES FINALES



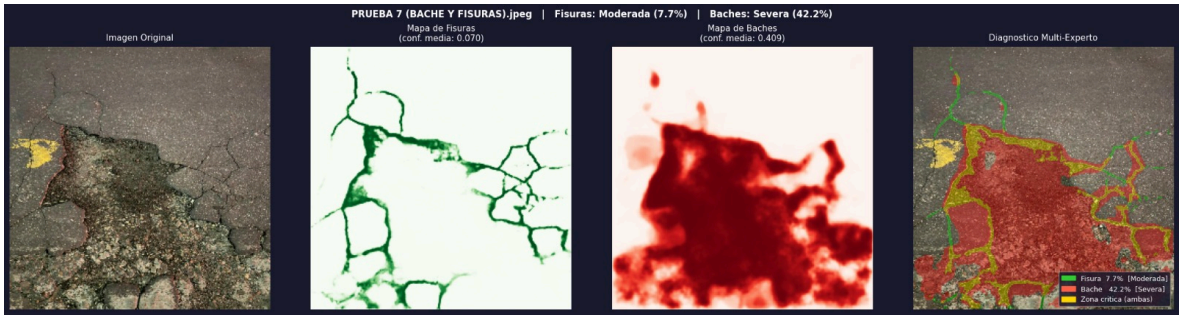
Fotografía tomada por el autor



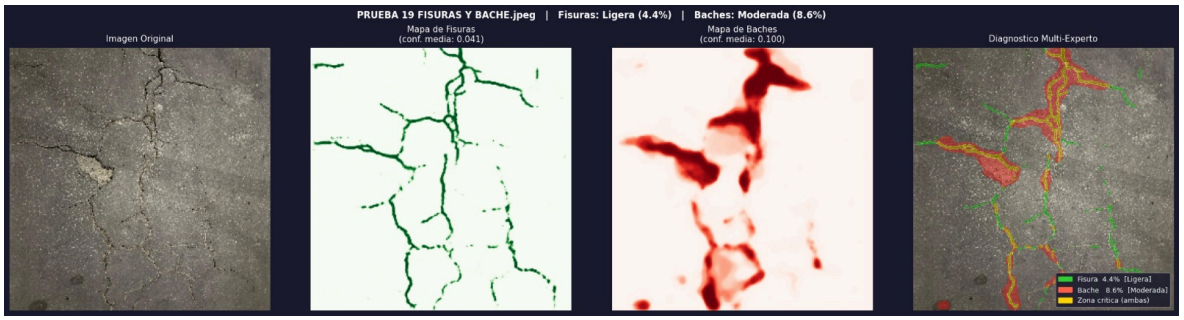
Fotografía tomada por el autor



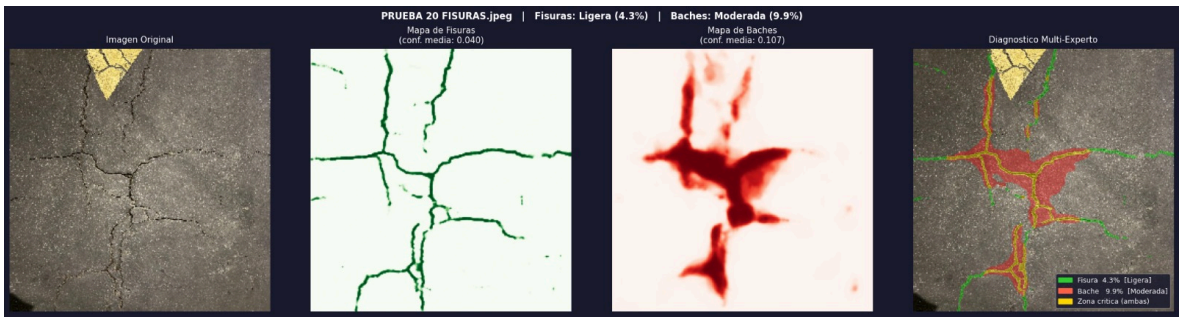
Fotografía tomada por el autor



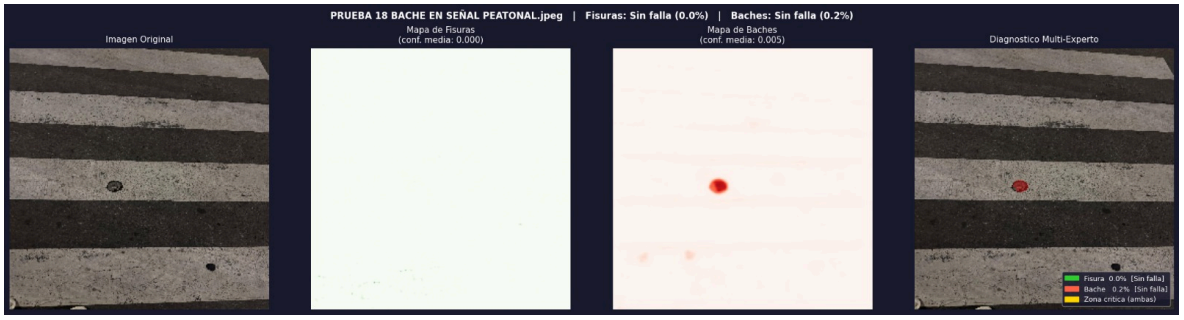
Fotografía tomada por el autor



Fotografía tomada por el autor



Fotografía tomada por el autor



Fotografía tomada por el autor

Tabla 10. Tabla de resultados obtenidos con imágenes reales. (Elaboración propia, 2026).

6.3 Formulación Matemática del Sistema Multi-Experto

En este apartado del trabajo de grado se busca sintetizar toda esta parte técnica de programación en términos matemáticos, así que el problema de segmentación de fallas viales se define como una función de mapeo desde un espacio de imágenes a un espacio de máscaras de probabilidad.

6.3.1 Definición del espacio para entrada y salida del modelo

Sea una imagen de entrada representada como un tensor $X \in R^{H \times W \times C}$ $donde\ H\ \ y\ W$ son la altura y anchura en píxeles (256 en este sistema), y C es el número de canales (1 para escala de grises, 3 para RGB). El objetivo del sistema es producir una máscara binaria de predicción $\hat{Y} \in \{0, 1\}^{H\,xW}$.

Con el objetivo de abordar la variabilidad morfológica de las fallas, se define un sistema multi-experto compuesto por dos funciones de aproximación (redes neuronales) independientes:

- El experto en fisuras: $\hat{P}_f = E_f\left(X_{gray}; \Theta_f\right)$

- El experto en baches: $\hat{P}_b = E_b\left(X_{rgb}; \Theta_b\right)$

Siendo Θ_f y Θ_b representaciones de los pesos entrenables de las arquitecturas Attention U-Net y U-Net-VGG 16, respectivamente, y $\hat{P}$ Es la matriz de probabilidades continuas en el rango [0,1].

6.3.2 Optimización y Funciones de Pérdida

Para el aprendizaje de los parámetros Θ se realiza minimizando funciones de pérdida diseñadas específicamente para lidiar con el desbalance de clases entre el pavimento sano y las fallas que hay en las imágenes.

Para el especialista en fisuras, se formula una pérdida combinada $L_{fisuras}$ que una la focal Loss (esto para mitigar el desbalanceo del fondo) y el Tversky Loss (el encargado de penalizar la omisión de grietas)

$$L_{fisuras} = L_{Focal}(p_t) + L_{Tversky}(\alpha, \beta)$$

Donde el focal Loss introduce un factor de modulación a la entropía cruzada estándar:

$$L_{Focal}(p_t) = -\alpha_f(1 - p_t)^{\gamma} \log log\,(p_t)$$

Y la Tversky Loss se define en términos de Verdades Positivos (TP), Falsos positivos (FP) y falsos negativos (FN):

$$L_{Tversky}(\alpha, \beta) = 1 - \frac{TP}{TP + \alpha \cdot FN + \beta \cdot FP}$$

Acá se usó el α = 0.7 y β = 0.3 para penalizar fuertemente los falsos negativos.

6.3.3. Inferencia con aumento de datos en tiempo de prueba (TTA)

Para mejorar la robustez de la inferencia sin alterar los parámetros Θ, se implementó una técnica de Test-Time Augmentation (TTA). Sea $T = \{T_1, T_2, ..., T_N\}$un conjunto de transformaciones geométricas (rotaciones y reflexiones espaciales). La matriz de probabilidad final $\hat{P}_{TTA}$ se calcula como la esperanza matemática de las predicciones sobre el espacio transformado:

$$\hat{P}_{TTA} = \frac{1}{N}\sum_{i=1}^{N} T_i^{-1}\left(E\left(T_i(X); \Theta\right)\right)$$

Donde T_i^{-1} Es la transformación inversa que realinea la predicción a la geometría original de la imagen.

6.3.4. Umbralización Dinámica y Fusión Multi-Experto

La binarización de las matrices de probabilidad continua no asume un umbral ingenuo de 0.5. En su lugar, el sistema define umbrales óptimos τ_f y τ_b mediante la maximización de la métrica de Intersección sobre Unión (IoU) en el conjunto de validación:

$$\tau^* = \arg arg\left(\frac{TP_\tau}{TP_\tau + FP_\tau + FN_\tau}\right)$$

Finalmente, la matriz booleana de diagnóstico global se obtiene aplicando la función indicadora *I* sobre los mapas de probabilidad de cada experto:

$$M_{fisuras} = I\left(\hat{P}_f > \tau_f^*\right)$$

$$M_{baches} = I\left(\hat{P}_b > \tau_b^*\right)$$

La fusión de estos espacios permite calcular el porcentaje de afectación vial A de manera aislada o conjunta para determinar el nivel de severidad:

$$A = \frac{\sum_{i,j} M_{fisuras} \cup M_{baches}}{H \times W} \times 100$$

Todo esto fue importante para entender el trasfondo matemático del modelo planteado en este trabajo de grado.

7. Conclusiones

Este trabajo de grado logró desarrollar e implementar un sistema funcional y generalizable de diagnóstico automatizado de fallas en pavimentos flexibles urbanos, dando cumplimiento a los objetivos propuestos y aportando evidencia cuantitativa, cualitativa y matemática sobre la viabilidad del enfoque Multi-Experto basado en segmentación semántica profunda. Los resultados obtenidos no solo validan las decisiones de diseño tomadas a lo largo del proceso, sino que establecen una base técnica concreta para la automatización del diagnóstico vial en ciudades como Bogotá, cuya malla vial requiere inversiones cercanas a los 11 billones de pesos para su rehabilitación integral.

Sobre la decisión arquitectónica central: el enfoque Multi-Experto

La elección de desacoplar el problema en dos arquitecturas especializadas, en lugar de mantener un modelo multiclase unificado, fue la decisión de diseño más determinante del trabajo y la que mayor impacto tuvo sobre los resultados. Los experimentos preliminares demostraron que una única red U-Net entrenada simultáneamente sobre fisuras y baches sufría de confusión de clases severa, produciendo un IoU de apenas 0.130 para agrietamiento, un resultado no esperado. Esta limitación no es un defecto del modelo en sí, sino una consecuencia estructural de intentar que una sola función de aproximación aprenda simultáneamente dos morfologías fundamentalmente distintas: las fisuras, estructuras lineales delgadas y finas que ocupan menos del 5% del área de cada imagen, y los baches, cavidades compactas de mayor superficie con bordes irregulares pero definidos.

Al desacoplar las arquitecturas y adaptar cada una a las características morfológicas de su clase objetivo, el IoU de fisuras pasó a 0.186 ± 0.007 en validación cruzada y el de baches alcanzó 0.403 ± 0.036, con un F1 de 0.611 ± 0.035. Estas mejoras, superiores al 50% para fisuras y al 10% para baches frente al modelo unificado, confirman de manera empírica que las fallas viales con morfologías estructuralmente distintas requieren tratamientos computacionales diferenciados. Esta conclusión tiene implicaciones directas para el diseño de futuros sistemas de inspección: no basta con incrementar la capacidad del modelo o la cantidad de datos si la arquitectura no está diseñada para la heterogeneidad morfológica del problema.

Sobre la gestión del desbalance de clases y la escasez de datos

Uno de los retos más significativos del trabajo fue la escasez de datos etiquetados para la clase de baches (710 imágenes) y el desbalance extremo de píxeles presente en ambas clases, donde el pavimento sano representa aproximadamente el 95% del área de cada imagen. La combinación de tres estrategias complementarias permitió mitigar este problema de manera efectiva: Transfer Learning desde el dataset HuggingFace CrackSeg (7.732 pares imagen-máscara), Data Augmentation con 8 transformaciones geométricas sincronizadas entre

imagen y máscara que expandieron el conjunto de baches de 568 a más de 8.500 iteraciones aumentadas, y funciones de pérdida asimétricas diseñadas específicamente para el dominio del problema.

La elección de $\alpha = 0.7$ en la función de Tversky Loss para el especialista en fisuras no respondió a una decisión arbitraria de hiper parámetro, sino a un criterio de riesgo explícito propio del dominio de inspección vial. En este contexto, no detectar una fisura existente tiene consecuencias estructurales y económicas considerablemente más graves que generar una falsa alarma que derive en una inspección innecesaria: una grieta capilar no atendida permite la infiltración de agua, acelera el deterioro estructural y puede evolucionar a un bache que implica costos de reparación varias veces superiores. Esta lógica de asimetría de costos, formalizada en la función de pérdida, conecta directamente la ingeniería de datos con la realidad operativa del mantenimiento vial, siendo uno de los aportes conceptuales más claros del trabajo.

Sobre la robustez y generalización del sistema

La validación cruzada de $K = 5$ folds proporciona la evidencia más sólida de que los modelos aprendieron patrones generalizables y no sobre ajustaron al conjunto de entrenamiento. El especialista en fisuras presentó una desviación estándar de IoU de apenas ± 0.007 a lo largo de los cinco folds, con un rango total de 0.019 entre el valor mínimo y máximo, lo que indica una estabilidad excepcional para un problema con desbalance tan extremo de clases. El especialista en baches alcanzó un IoU medio de 0.403 ± 0.036 y un F1 de 0.611 ± 0.035, con mayor varianza atribuible al menor volumen de datos disponibles (710 imágenes frente a 2.852 de fisuras).

Ambos valores se sitúan dentro de los rangos reportados en la literatura especializada de segmentación vial: IoU entre 0.15 y 0.35 para fisuras en condiciones de alto desbalance, e IoU entre 0.35 y 0.65 para detección de baches. Que el sistema propuesto alcance estos rangos con un dataset de tamaño moderado, sin infraestructura de captura especializada y con imágenes tomadas

en condiciones reales del entorno bogotano, válida la solidez metodológica del enfoque y sugiere que los resultados son escalables con mayor disponibilidad de datos locales.

La prueba de generalización más contundente no es estadística sino visual: el sistema fue evaluado con fotografías tomadas por el autor en calles de Bogotá, bajo condiciones variables de iluminación, ángulo y calidad de captura, y en imágenes obtenidas de fuentes externas nunca vistas durante el entrenamiento. En estos escenarios, el sistema demostró capacidad para detectar simultáneamente ambos tipos de falla sin solapamiento erróneo, delimitar morfológicamente los bordes del daño, ignorar objetos intrusivos dentro de un bache y producir un diagnóstico de severidad interpretable. Esta capacidad de operar con equipos convencionales sin condiciones controladas de captura es un requisito indispensable para cualquier implementación práctica en el contexto de las entidades de mantenimiento vial de Bogotá.Aunque tuvo ciertas limitaciones los resultados fueron maravillosos.

Sobre las limitaciones identificadas y su significado metodológico

El sistema presentó vulnerabilidades ante condiciones específicas de captura que deben comunicarse con claridad al usuario final del diagnóstico. La presencia de agua, escombros u objetos ajenos dentro del área de un bache interrumpe la continuidad de los gradientes de textura que el encoder utiliza para la extracción de características, generando segmentaciones parciales o falsos negativos que subestiman el área total del daño estructural. Asimismo, sombras proyectadas por el operario durante la captura y elementos con contornos oscuros en el borde del encuadre fueron clasificados erróneamente como fallas, debido a su similitud visual con la morfología de un bache.

Por su parte, el especialista en fisuras demostró dificultad para distinguir entre una fisura activa y una intervención previa de sellado, dado que ambas comparten características visuales de contraste y bordes definidos. Esta limitación no invalida el sistema ni sus métricas de desempeño,

sino que delimita con precisión los escenarios donde su confiabilidad disminuye, información valiosa para establecer protocolos de uso responsable. Desde una perspectiva metodológica, la identificación y documentación de estas vulnerabilidades constituye en sí misma un resultado del trabajo: sin ella, cualquier implementación operativa del sistema carecería del criterio necesario para interpretar correctamente las predicciones del modelo en campo.

Sobre el impacto potencial y la relevancia para Bogotá

El sistema desarrollado representa un primer paso concreto y técnicamente fundamentado hacia la automatización del diagnóstico vial preventivo en Bogotá. Al generar un diagnóstico visual interpretable con clasificación de severidad en cuatro niveles operativos, cuantificación del porcentaje de área afectada y capacidad de procesar imágenes capturadas con equipos convencionales sin infraestructura especializada, el sistema propone una alternativa escalable a las inspecciones manuales actualmente empleadas por entidades como el IDU y la UMV.

La arquitectura Multi-Experto presentada es directamente compatible con los sistemas de reporte ciudadano existentes en la ciudad, como la plataforma web de la UMV y el Sistema Distrital de Quejas y Soluciones. Una integración futura de este sistema con dichas plataformas permitiría enriquecer automáticamente los reportes ciudadanos con un diagnóstico técnico cuantificado del daño, priorizando las intervenciones según la severidad detectada y reduciendo la dependencia de la subjetividad inherente a las inspecciones visuales manuales. En un contexto donde el costo de rehabilitar la malla vial de Bogotá supera los 11 billones de pesos, la detección temprana y precisa de fallas en etapa incipiente tiene un valor económico y social directo que justifica plenamente la inversión en sistemas automatizados de este tipo.

Líneas de trabajo futuro

Lo primero que debe tenerse en cuenta como línea para futuros trabajos es que, aunque se obtuvieron resultados positivos y llamativos, durante este estudio se presentaron limitaciones físicas y geométricas inherentes al procesamiento de imágenes bidimensionales. En este contexto, la clasificación ad hoc de severidad propuesta, basada en la proporción porcentual de píxeles afectados respecto al total de la imagen, se encuentra condicionada por dos sesgos algorítmicos y operativos fundamentales, los cuales impiden considerar estas mediciones como representaciones absolutas del volumen físico real del daño.

Primero, se presentó un sesgo asociado a la compresión espacial (redimensionamiento). La arquitectura U-Net empleada requiere la estandarización de los tensores de entrada a una resolución fija de 256 × 256 píxeles. Dado que las fotografías originales del entorno vial poseen resoluciones significativamente mayores, generalmente superiores a 1920 × 1080 píxeles, el proceso de reducción espacial introduce una pérdida de fidelidad topológica.

En el caso específico de las fisuras en pavimentos, que en alta resolución pueden ocupar apenas unos pocos píxeles de ancho, el algoritmo de interpolación tiende a difuminar o incluso eliminar estos micro patrones al promediarlos con el asfalto circundante. Adicionalmente, el hecho de adaptar imágenes con relaciones de aspecto rectangulares a una matriz cuadrada modifica la excentricidad geométrica de las fallas, alterando el área relativa que el modelo termina segmentando.

El segundo sesgo limitante se encuentra en la metodología de adquisición de datos. Las imágenes empleadas fueron obtenidas a partir de datasets externos, los cuales presentaban limitaciones en cuanto a la diversidad de perspectivas y otros factores relevantes para la captura de los defectos. Esta situación se debe, principalmente, a la escasez de conjuntos de datos disponibles que contengan imágenes de fisuras y baches en pavimentos.

Como consecuencia, el área en píxeles que ocupa un bache dentro de una fotografía es inversamente proporcional al cuadrado de la distancia entre la lente de la cámara y el pavimento. De esta manera, una anomalía de 30 centímetros de diámetro fotografiada a un metro de distancia

abarcará un porcentaje considerablemente mayor del encuadre que ese mismo daño capturado a tres metros.

Por lo tanto, el porcentaje de severidad calculado por el algoritmo desarrollado en este trabajo de grado representa una magnitud escalar de carácter puramente topocéntrico, es decir, relativa al plano focal del sensor, y no una medición volumétrica ni kilométrica correspondiente al mundo real. En consecuencia, para que el sistema pueda realizar inferencias de ingeniería civil con un mayor nivel de precisión absoluta, los trabajos futuros deberían contemplar la integración de cámaras fijas instaladas a una altura y un ángulo estrictamente estandarizados sobre el chasis de vehículos de inspección, o bien la implementación de otras alternativas que permitan mitigar esta limitación.

Los resultados de este trabajo sugieren tres líneas de desarrollo prioritarias. La primera es la inclusión sistemática de ruidos ambientales, como sombras, agua, objetos intrusivos y pavimento mojado, en las estrategias de Data Augmentation, para robustecer la generalización del modelo ante las condiciones adversas de captura más frecuentes en el entorno vial urbano. La segunda es la construcción de un dataset propio de baches y fisuras capturado específicamente en vías de Bogotá, con anotaciones a nivel de píxel validadas por ingenieros civiles especializados en patología vial; este dataset local permitiría reducir la varianza observada en la validación cruzada del especialista en baches y mejorar la adaptación al dominio local. La tercera, y de mayor impacto operativo, es la integración del sistema en una aplicación móvil o plataforma web con soporte de georreferenciación, que permita conectar el diagnóstico automático con los sistemas distritales de mantenimiento vial, habilitando la priorización automatizada de intervenciones y cerrando el ciclo completo entre la detección temprana de la falla y su atención por parte de las entidades responsables.

En síntesis, este trabajo demuestra que es técnicamente viable construir un sistema de segmentación semántica Multi-Experto para la detección automatizada de fallas en pavimentos flexibles, con métricas coherentes con el estado del arte, capacidad de generalización demostrada

sobre imágenes reales del entorno bogotano, y una arquitectura diseñada desde principios del dominio de la ingeniería civil. Los resultados obtenidos no son el punto final de este problema, sino la base metodológica sobre la cual construir sistemas de mayor escala, mayor cobertura de datos locales y mayor integración con las plataformas de gestión vial de la ciudad.

8. Bibliografía

Frapial. (2022, diciembre 3). Factores que influyen en el deterioro del asfalto. **https://www.frapial.com/factores-que-influyen-en-el-deterioro-del-asfalto/**

NAC Supply, Inc. (2019, agosto 13). *How different weather affects asphalt pavement*.

Guzmán, Z. (2025, julio 22). Maratón Tapahuecos: IDU interviene malla vial de todas las localidades de la ciudad. Caracol Radio. https://caracol.com.co/2025/07/22/maraton-tapahuecos-idu-interviene-malla-vial-de-todas-las-localidades-de-la-ciudad/

Yoder, E. J., & Witczak, M. W. (1975). Principles of pavement design (2nd ed.). John Wiley & Sons.

Instituto Nacional de Vías (INVIAS). (2013). Especificaciones generales de construcción de carreteras.

EADIC. (2015, febrero 23). *Pavimento rígido vs pavimento flexible en Latinoamérica*. https://eadic.com/blog/entrada/pavimento-rigido-vs-pavimento-flexible-en-latinoamerica/

Fripal. (2022). *Pavimento rígido*. https://www.frapial.com/

Ekimak Engineering. (2023). *Esfuerzo en pavimento flexible y rígido*.

Giordani, C. (s.f.). *Capas para la construcción de la vía*.

Instituto Nacional de Vías (INVIAS). (2012). *Normas de ensayo de materiales para carreteras (Secciones 700–800)*.

Correa, J. C., & Palacio, C. A. (2017). *Evaluación del deterioro superficial en pavimentos flexibles mediante técnicas de procesamiento digital de imágenes*. *Revista Facultad de Ingeniería*, (82), junio de 2017.

Instituto de Desarrollo Urbano (IDU). (2023). *Estado de la malla vial de Bogotá*.

Noticias Caracol. (2024, junio 18). *¿Con qué tipo de asfalto se pavimentan las calles de Bogotá?* [Video]. YouTube. https://www.youtube.com/watch?v=E3T2iN6SQow.

Alcaldía de Bogotá. (2023). *Sistema Distrital de Quejas y Soluciones: reporte de huecos.*

Federal Highway Administration. (2003). *Distress identification manual for the long-term pavement performance program (FHWA-RD-03-031)*. U.S. Department of Transportation. https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/reports/03031/03031.pdf

PMS. (2021). Common causes of pavement failure. Pavement Management Services. https://www.pavement.com.au/news/6-causes-of-pavement-failure

Wiley. (2008). The Wiley Encyclopedia of Management (2nd ed.). Wiley.

McGuire, S. (2020, 26 junio). *Potholes—An Expensive and Recurring Problem for Cities*. Giatec Scientific Inc.

Ramírez, M. (2024, 13 julio). El país de los baches: qué puede hacer el nuevo Gobierno laborista contra la plaga de las calles del Reino Unido. *ElDiario.es*. eldiario.es/internacional/pais-baches-nuevo-gobierno-laborista-plaga-calles-reino-unido_1_11518101.html#:~:text=Por%20los%20datos%20de%20una,comparación%20%20con%%2020el%20año%20anterior.&text=Arreglar%20los%20baches%20se%20ha,millón%2%200de%20baches%20al%20año

Hernández, C. (2025, 10 de agosto). ¿Cuánto dinero se requiere para hacer el mantenimiento de toda la malla vial de Bogotá? Portafolio.

Alcaldía de Bogotá. (2024, 18 de diciembre). *Intervención de huecos en Bogotá: IDU y grupos de moteros se unen.*

Garzón, D. (2025, 23 julio). 'Maratón Tapahuecos' en las calles de Bogotá: estos son los sectores que intervendrá el distrito. Infobae. *https://www.infobae.com/colombia/2025/07/23/maraton*https://www.infobae.com/colombia/2025/07/23/maraton-tapahuecos-en-las-calles-de-bogota-estos-son-los-sectores-que-intervendra-el-distrito/*tapahuecosen-las-calles-de-bogota-estos-son-los-sectores-que-intervendra-el-distrito/*

Noticias Caracol. (2021, 12 noviembre). Conozca la nueva app para reportar huecos en Bogotá [Vídeo]. YouTube. https://www.youtube.com/watch?v=2zHkQxgAdb8

Alcaldía de Bogotá. Sánchez, G. R. (2023, 2 de mayo). Así puedes reportar los huecos en las vías de Bogotá para que sean tapados. Bogotá.gov.co. https://bogota.gov.co/mi-ciudad/movilidad/aprendecomo-reportar-los-huecos-en-bogota-para-que-sean-tapados

Guía Diseño Vías Urbanas — IDU. (s. f.). Guía vías urbanas [Documento en Scribd]. Scribd. https://es.scribd.com/document/797655302/GUIA-VIAS-URBANAS-IDU

Instituto de Desarrollo Urbano (IDU). (2016). Especificaciones técnicas generales de materiales. Bogotá: IDU.

Instituto Nacional de Vías (INVÍAS). (2013, 11 de abril). *Manual de drenaje para carreteras*. INVÍAS.

Concejo de Bogotá. (2020). *DTs Geometría* [PDF]. Bogotá, Colombia: Concejo de Bogotá. Recuperado de https://concejodebogota.gov.co/cbogota/site/artic/20200613/asocfile/20200613074940/dts_geometria.pdf

Sánchez, G. R. (2024, 2 de mayo). *Así es la producción de asfalto en Bogotá para tapar huecos en vías.* Bogotá.gov.co.

Noticias caracol. (2024). *¿Con qué tipo de asfalto se pavimentan las calles de Bogotá?* [Video]. YouTube. https://www.youtube.com/watch?v=E3T2iN6SQow

Ma, N., Fan, J., Wang, W., Wu, J., Yu, J., Xie, L., & Fan, R. (2022). *Computer vision for road imaging and pothole detection: A state-of-the-art review of systems and algorithms*. *Transportation Safety and Environment, 4*(4). https://doi.org/10.1093/tse/tdac026

Satti, S. K., Rajareddy, G. N. V., Mishra, K., & Gandomi, A. H. (2024). *Potholes and traffic signs detection by classifier with vision transformers*. *Scientific Reports, 14*, Article 2215. https://doi.org/10.1038/s41598-024-52426-4

Ríos Cotazo, N. X., Bacca Cortés, B., Caicedo Bravo, E., Orobio Quiñónez, A. (2020). *Revisión de métodos para la clasificación de fallas superficiales en pavimentos flexibles*. Ciencia e Ingeniería Neogranadina, 30(2), 109-127. https://doi.org/10.18359/RCIN.4385

Galindo-López, M. de J., Salazar-Segundo, A., & Hernández-Aguilar, J. A. (2024). *Detección automática de averías en el pavimento mediante visión por computadora y cómputo móvil.* Research in Computing Science, 153(7), 163-175. https://rcs.cic.ipn.mx/2024_153_7/Deteccion%20automatica%20de%20averias%20en%20el%20pavimento%20mediante%20vision%20por%20computadora.pdf

UNSCH. (2022). *GRUPO N° 06 - EVALUACIÓN DE DAÑOS EN PAVIMENTOS FLEXIBLES USANDO FOTOGRAMETRÍA TERRESTRES* [Video]. YouTube. https://www.youtube.com/watch?v=9XUxkdAWoH4

IFCA News. (2024). *Un investigador de la UC crea un modelo de IA capaz de detectar carreteras deterioradas* [Video]. YouTube. https://www.youtube.com/watch?v=RTrCOD1dyg4

Peña Urrea, C. C., & Cáceres Contreras, L. F. (2021). *Caracterización de una muestra de pavimento asfáltico reciclado (RAP), extraído de la Autopista Sur a la altura de los municipios de Soacha y Sibaté Cundinamarca, con el fin de determinar su uso en capas granulares no tratadas en una estructura de pavimento* [Tesina]. Universidad Distrital Francisco José de Caldas. https://repository.udistrital.edu.co/items/21056872-9818-47a3-bb3c-6d499b1ce10b

Alcaldía Mayor de Bogotá – Secretaría de Tránsito y Transporte. (2005). *Manual de Planeación y Diseño para la Administración del Tránsito y el Transporte. Tomo I: Marco Conceptual* (2ª ed.). Bogotá: Alcaldía Mayor de Bogotá. https://www.movilidadbogota.gov.co/web/sites/default/files/tomo_i.pdf Scribd+2Google Libros+2

Instituto de Desarrollo Urbano (IDU). (2016). *Especificaciones técnicas generales de materiales*. Bogotá: IDU. Recuperado de https://www.idu.gov.co/page/especificaciones-tecnicas-generaleshttps://www.idu.gov.co/page/especificaciones-tecnicas-generales-de-materiales?utm_source=chatgpt.comdemateriales

Alcaldía de Bogotá – Secretaría de Movilidad. (2024). *Plan de mantenimiento de malla vial: 53.2 kmcarril intervenidos en menos de un mes*. Bogotá.gov.co. Recuperado de https://bogota.gov.co/miciudad/movilidad/plan-de-mantenimiento-malla-vial-en-menos-de-un-meshttps://bogota.gov.co/mi-ciudad/movilidad/plan-de-mantenimiento-malla-vial-en-menos-de-un-mes-mas-de-50-kmmas-de-50-km

Ministerio de Obras Públicas y Comunicaciones (MOPC), República Dominicana. (2016). *Identificación de fallas en pavimentos y técnicas de reparación (Catálogo de fallas)*. https://www.mopc.gob.do/media/2335/sistema-identifici%C3%B3n-fallas.pdf mopc.gob.do

Alcaldía Mayor de Bogotá – Secretaría de Movilidad. (2015). *UMV ha tapado 39 mil 161 huecos de la malla vial local*. Bogotá.gov.co. Recuperado de https://bogota.gov.co/mi-ciudad/movilidad/umvhttps://bogota.gov.co/mi-ciudad/movilidad/umv-ha-tapado-39-mil-161-huecos-de-la-malla-vial-localhatapado-39-mil-161-huecos-de-la-malla-vial-local

University transportation research center. (2017). *Innovative techniques for maintenance repair and reconstruction (mrr) of asphalt roadways*. utrc. https://www.utrc2.org/sites/default/files/Final-ReportInnovative-Techniques-for-Maintenance-Repair-Reconstruction.pdf

Amado Marín, j. f. (2015). *Análisis del sistema de reparación de pavimentos flexibles por inyección neumática de mezclas asfálticas en frío, tecnología velocity patching* [tesis de maestría, universidad distrital francisco josé de caldas]. repositorio institucional distrital.

https://repository.udistrital.edu.co/server/api/core/bitstreams/240ac05c-ef7f-40e9https://repository.udistrital.edu.co/server/api/core/bitstreams/240ac05c-ef7f-40e9-8cec-e6d170661d0a/content8cece6d170661d0a/content

Huertas Becerra, R., & Cely Herazo, R. (2013). *Memoria técnica sobre la rehabilitación de 3400 m2 de vías locales de acceso a bodegas ubicadas en la carrera 7 con calle 200 costado occidental en Bogotá*
[trabajo académico, universidad militar nueva granada]. repositorio umng. https://repository.umng.edu.co/server/api/core/bitstreams/007b4cae-f7a5-4785https://repository.umng.edu.co/server/api/core/bitstreams/007b4cae-f7a5-4785-b5f9-efeacf038617/contentb5f9efeacf038617/content

Ríos Cotazo, N. X., Bacca Cortés, B., Caicedo Bravo, E., Orobio Quiñónez, a. (2020). *Revisión de métodos para la clasificación de fallas superficiales en pavimentos flexibles*. ciencia e ingeniería neogranadina, 30(2), 109–127. https://revistas.umng.edu.co/index.php/rcin/article/view/4385

Escobar-Arenas, J. P., Flórez-Pareja, L. D., & Fernandez-Mc Cann, D. S. (2023).Estimación *de irregularidades en pavimentos mediante técnicas de procesamiento digital de imágenes*. Revista ingeniería solidaria, 19(1), 1866. https://www.redalyc.org/journal/6078/607874723002/html

Suárez, r. j. (2023). *Sistema para analizar señales de sensores en celulares que permitan identificar y analizar la mala calidad del pavimento en las vías* [tesis de maestría, Pontificia Universidad Javeriana]. api dspace.
https://apidspace.javeriana.edu.co/server/api/core/bitstreams/28e1b2c6-ad6b-45cchttps://apidspace.javeriana.edu.co/server/api/core/bitstreams/28e1b2c6-ad6b-45cc-975d-a475bc75a58b/content975da475bc75a58b/content

Qingbo, Z. (2016). Pavement crack detection algorithm based on image processing analysis. *2016 8th international conference on intelligent human-machine systems and cybernetics (ihmsc)*, 15–18. ieee.
https://doi.org/10.1109/IHMSC.2016.96

Tello-Cifuentes, L., Aguirre-Sánchez, M., Díaz-Paz, J. P., & Hernández, F. (2021). Evaluación de daños en pavimento flexible usando fotogrametría terrestre y redes neuronales. *tecnológicas, 24*(50), 1866.
https://doi.org/10.22430/22565337.1686

Buchinger, D., & Silva, A. G. (2014). Anomalies detection in asphalt pavements: a morphological image processing approach. Revista Brasileira de Computação Aplicada, 6(1), 121-129. http://dx.doi.org/10.5335/rbca.2014.3661

Escalona, U., Arce, F., Zamora, E., & Sossa, H. (2019). Fully Convolutional Networks for automatic pavement crack segmentation. Computación y Sistemas, 23(2), 451-460. https://doi.org/10.13053/CyS-23-2-3047

Miranda-Argüello, F., & Aguiar-Moya, J. P. (2019, enero). *Mezclas asfálticas con RAP: Pavimentos asfálticos reciclados* (Vol. 10, N.º 1). Lanamme-UCR. https://www.lanamme.ucr.ac.cr/repositorio/bitstream/handle/50625112500/1421/Boletin%201%20%20Mezclas%20asf%C3%A1lticas%20con%20RAP.pdf?sequence=1

Codificando bits. (2024, abril 8). *Tutorial: SEGMENTACIÓN DE IMÁGENES SATELITALES con Redes Convolucionales (U-Net)* [Video]. YouTube. https://www.youtube.com/watch?v=MSiVi9I8SFU

DataScientest. (2024, 26 Abr.). *U-NET: todo lo que tienes que saber sobre la red neuronal de Computer Vision*. https://datascientest.com/es/u-net-lo-que-tienes-que-saber

Nekouei, F. (2025, julio). *Image Segmentation | Pothole Detection using U-Net and ResNet34* [Video]. YouTube. https://www.youtube.com/watch?v=KCuax2yLlZI

Torres Loaiza, J. E. (2004). *El rigor: garantía para los periodistas, credibilidad para los medios y esperanza para los ciudadanos*. http://www.redalyc.org/pdf/649/64901105.pdf

Anexos: Repositorio de Datos y Modelos

En este apartado al tener conjuntos de datos y modelos muy pesados se decidió tomar y manejar todo desde un drive, por lo que a continuación se compartirán tanto los datos, los modelos y el código final que se usó:

- **Datos principales:**

  https://drive.google.com/drive/folders/1vjVs_G75CP6UzmxQzzjYBcwrXT9bMhnl?usp=sharing
- **Datos de Hugging Face:**

  https://drive.google.com/drive/folders/11M7Csk9AxubXiQM_pK_gjM3YP6Gb7clL?usp=sharing
- **Modelos:**

  https://drive.google.com/drive/folders/1-cQ5Y5LjYSo1exKrUch45EgbGhDyekXa?usp=sharing
- **Código final:**

  https://colab.research.google.com/drive/1CTJmOj-ItZ8hSjMW-epwRmHnDXfc9O-v?usp=sharing
- **Codigo para pruebas:**

  https://colab.research.google.com/drive/1rklfITjpsSZa3vZtNSKxG4DA180ZZDZi?usp=sharing