



**REPÚBLICA DE PANAMÁ  
UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA  
FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA**

**INGENIERÍA EN REDES DE COMUNICACIONES CON ÉNFASIS EN SEGURIDAD**

**PROTOTIPO CON VISIÓN COMPUTACIONAL EXPLICABLE-SEMÁNTICO PARA  
DETECCIÓN DE ACUMULACIÓN DE RESIDUOS EN LAS CALLES DE SAN  
MIGUELITO, CIUDAD DE PANAMÁ**

**PROYECTO DE TRABAJO PARA OPTAR AL GRADO DE LICENCIADO EN  
INGENIERÍA EN REDES DE COMUNICACIONES CON ÉNFASIS EN SEGURIDAD**

**Tutor: José Rivera Castro**

**Autor: Jorge Andres Gregory Guette**

Ciudad de Panamá, 15 de octubre de 2025



**REPÚBLICA DE PANAMÁ  
UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA  
FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA**

**INGENIERÍA EN REDES DE COMUNICACIONES CON ÉNFASIS EN SEGURIDAD**

**PROTOTIPO CON VISIÓN COMPUTACIONAL EXPLICABLE-SEMÁNTICO PARA  
DETECCIÓN DE ACUMULACIÓN DE RESIDUOS EN LAS CALLES DE SAN  
MIGUELITO, CIUDAD DE PANAMÁ**

**PROYECTO DE TRABAJO PARA OPTAR AL GRADO DE LICENCIADO EN  
INGENIERÍA EN REDES DE COMUNICACIONES CON ÉNFASIS EN SEGURIDAD**

**Autor: Jorge Andres Gregory Guette**

Ciudad de Panamá, 15 de octubre de 2025



Ciudad de Panamá, 9 de octubre de 2025

Profesor(a)

Héctor Mazurkiewicz

Coordinador del Comité de Titulación de Estudios de Licenciatura.

Presente.

En mi carácter de Tutor del Trabajo de Grado presentado por el estudiante Jorge Andres Gregory Guette, documento de pasaporte N.º AS792631, para optar al grado de Licenciatura en Ingeniería en Redes de Comunicaciones con Énfasis en Seguridad considero que el trabajo: reúne los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte del Jurado examinador que se designe.

Atentamente,

---

José Rivera Castro

Documento de identidad: cédula No. 3-707-28

Línea de Investigación: visión por computadora.



**UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA**  
**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA**

**INFORME DE ACTIVIDADES DE TUTORÍA OPCIÓN DE TITULACIÓN II**

Estudiante: Jorge Andres Gregory Guette. Pasaporte No. AS792731, tutor: Prof. José Rivera Castro; cédula de identidad No. 3-707-28; correo electrónico del participante: [jorgegregory@hotmail.com](mailto:jorgegregory@hotmail.com); celular No. 6203-4497. Título tentativo del trabajo de grado (TG): **PROTOTIPO CON VISIÓN COMPUTACIONAL EXPLICABLE-SEMÁNTICO PARA DETECCIÓN DE ACUMULACIÓN DE RESIDUOS EN LAS CALLES DE SAN MIGUELITO, CIUDAD DE PANAMÁ**. Línea de Investigación: visión por computadora.


Sesión	Fecha	Hora Reunión.	Aspecto Tratado	Observación
1.	Miércoles 6 de agosto	07:30 PM	Revisiones del capítulo 1	Correcciones ortográficas
2.	Domingo 10 de agosto	08:30 pm	Revisiones anteproyecto.	Correcciones en justificación y planteamiento del problema
3.	Miércoles 20 de agosto	11:30 am	revisiones objetivas, marcos teóricos	Mejoras en los objetivos y agregar más antecedentes
4.	Viernes 22 de agosto	09:30 pm	Correcciones generales	Correcciones generales

<b>5.</b>	Miércoles 23 de septiembre	07:00 pm	Correcciones generales	Correcciones generales
<b>6.</b>	Sábado 4 de octubre	09:00 am	Correcciones metodología	Correcciones metodología

Título definitivo: PROTOTIPO CON VISIÓN COMPUTACIONAL EXPLICABLE-SEMÁNTICO PARA DETECCIÓN DE ACUMULACIÓN DE RESIDUOS EN LAS CALLES DE SAN MIGUELITO, CIUDAD DE PANAMÁ.

Comentarios finales acerca de la investigación: Declaramos que las especificaciones anteriores representan el proceso de dirección del trabajo de grado arriba mencionado.

  
\_\_\_\_\_  
Firma

  
\_\_\_\_\_  
Firma

## **DEDICATORIA**

Quiero dedicar este logro a mi familia, en especial a mi abuela María Teresa por ser mi apoyo y lo más hermoso en mi vida que estará muy orgullosa de verme alcanzar un logro más en la vida, a mi esposa Iravik y mi madre Dolly Esther por estar siempre a mi lado brindándome su apoyo y las fuerzas para seguir adelante, demostrando que con esfuerzo y perseverancia podemos lograr escalar un peldaño más en la vida. También quiero dedicar este logro al Mg. José Rivera Castro, por el apoyo incondicional en la culminación de este proyecto. Le quiero dar gracias a mi Dios Jehová por permitirme finalizar este enorme reto, la cual se la dedico a mi familia, quienes han estado conmigo apoyándome durante todo este tiempo y a mi hermosa esposa Iravik mi compañera fiel, quien me apoya en todo lo que realizo, sin ellos no hubiese podido alcanzar este logro.

## **RECONOCIMIENTO**

Agradezco a Dios por permitir ser su instrumento para llevar a cabo este estudio y alcanzar una meta más en mi vida, por guiar me en cada paso con sabiduría e inteligencia para culminar con éxito esta etapa, que nos permitirá crecer más como profesionales y poder servir a la sociedad con nuestros conocimientos adquiridos. Agradecemos al claustro de profesores que nos acompañaron durante todos nuestros estudios, por su profesionalismo y magisterio a cada momento. En especial a nuestro Coordinador Dr. Héctor Mazurkiewicz por compartir sus conocimientos, por su apoyo y colaboración para la elaboración de este proyecto. A esta prestigiosa Universidad, que va a la vanguardia con mejor preparación de sus profesionales, que nos acogió y nos premió de conocimientos que nos permitirán aportar las mejores herramientas pedagógicas en formación de muchos hombres y mujeres que serán el futuro del país.

## ÍNDICE GENERAL

	Pág.
PORADA.....	1
PORADA INTERNA.....	2
CARTA DE APROBACIÓN DEL TUTOR.....	3
INFORME DE ACTIVIDADES DE TUTORÍA OPCIÓN DE TITULACIÓN II.....	4
DEDICATORIA.....	6
RECONOCIMIENTO.....	7
ÍNDICE GENERAL.....	8
ÍNDICE DE TABLAS.....	12
ÍNDICE DE FIGURAS.....	13
RESUMEN.....	15
<b>CAPÍTULOS</b>	
<b>INTRODUCCIÓN.....</b>	<b>17</b>
<b>PLANTEAMIENTO DEL PROBLEMA.....</b>	<b>18</b>
Justificación.....	19
Pregunta de Investigación.....	19
Objetivos.....	20
Objetivo General.....	20
Objetivos Específicos.....	20
Hipótesis.....	20
Alcance y Delimitación del Proyecto.....	21
Marco Teórico.....	22
Antecedentes.....	22
Acumulación de Residuos en las Calles.....	22
Visión Computacional.....	26
Visión Computacional para Detectar Residuos.....	28
Visión Computacional Explicable.....	31
Bases Teóricas.....	34
Visión Computacional.....	34

Tipos de Visión Computacional.....	35
Inteligencia Artificial Explicable.....	35
Definición.....	35
Inteligencia Artificial Responsable.....	36
Conceptos Fundamentales.....	36
Taxonomía de Inteligencia Artificial Explicable.....	37
Taxonomía General para Aprendizaje Automático.....	37
Taxonomía Específica para Aprendizaje Profundo.....	37
Oportunidades de la Inteligencia Artificial Explicable.....	39
Principales Desafíos.....	40
Modelos de Lenguaje de Gran Escala.....	41
Definición.....	41
Representación Semántica.....	41
Construcción de Teorías.....	42
Limitaciones.....	42
Text-to-SQL Basado en Modelos de Lenguaje de Gran Escala.....	42
Definición.....	42
Objetivo.....	43
Componentes Clave.....	43
Métricas de Evaluación.....	44
Fases en Proyectos de Inteligencia Artificial.....	44
Bases Legales.....	45
Definición de Residuo.....	45
Responsabilidades de la Autoridad de Aseo Urbano y Domiciliario.....	45
Términos básicos.....	46
<b>METODOLOGÍA.....</b>	51
Métodos de Investigación.....	51
Tipos de investigación.....	51
Variables.....	52
Operacionalización de Variables.....	53
Población y Muestra.....	53

Técnica e Instrumentos.....	55
Procedimientos.....	55
Construir el prototipo: Fase de Diseño y Entrenamiento.....	55
Selección de Explicabilidad.....	56
Obtención de Límites Administrativos.....	56
Historial de Imágenes.....	56
Captura de Imágenes Satelitales.....	57
Etiquetado y Preparación del Conjunto de Imágenes.....	57
Entrenamiento del Modelo.....	57
Diseño de la Aplicación Web.....	58
Contabilizar la Acumulación de Residuos: Fase de Diseño y Entrenamiento.....	58
Medir el rendimiento general del modelo: fase de evaluación de generalización.....	59
Evaluar la Fidelidad Explicativa-Semántica: Fase de Evaluación simulada.....	59
ANÁLISIS DE DATOS Y RESULTADOS.....	61
Análisis de Datos.....	61
Construir el Prototipo: Fase de Diseño y Entrenamiento.....	61
Contabilizar la Acumulación de Residuos: Fase de Diseño y Entrenamiento.....	62
Medir el Rendimiento General del Modelo: Fase de Evaluación de Generalización.....	62
Evaluar la Fidelidad Explicativa-Semántica: Fase de Evaluación Simulada.....	62
Resultados.....	63
Construir el prototipo: Fase de Diseño y entrenamiento.....	63
Contabilizar la Acumulación de Residuos: Fase de Diseño y Entrenamiento.....	64
Medir el Rendimiento General del Modelo: Fase de Evaluación de Generalización.....	66

Evaluar la Fidelidad Explicativa-Semántica: Fase de Evaluación	
Simulada.....	67
<b>CONCLUSIONES.....</b>	69
Construir el Prototipo: Fase de Diseño y	
Entrenamiento.....	69
Contabilizar la Acumulación de Residuos: Fase de Diseño y	
Entrenamiento.....	71
Medir el Rendimiento General del modelo: Fase de Evaluación de	
Generalización.....	72
Evaluar la Fidelidad Explicativa-Semántica: Fase de Evaluación	
Simulada.....	72
<b>LIMITACIONES.....</b>	75
<b>RECOMENDACIONES.....</b>	76
<b>REFERENCIAS.....</b>	77
<b>ANEXO A.....</b>	84
<b>ANEXO B.....</b>	90

## ÍNDICE DE TABLAS

<b>Tabla No.</b>		<b>Pág.</b>
A1	Resumen de Antecedentes y su Aporte	84
A2	Criterios para Aprendizaje Automático	87
A3	Resumen de Concepto y su Aporte	87
B1	Operacionalización de Variables	90
B2	Teselas de la Población por Corregimiento	92
B3	Distribución Primer Entrenamiento	93
B4	Distribución Segundo Entrenamiento	93
B5	Distribución Tercer Entrenamiento	94

## ÍNDICE DE FIGURAS

<b>Figura</b>		<b>Pág.</b>
<b>No.</b>		
B1	Distribución de Imágenes de Entrenamiento	94
B2	Ficha Observación y sus Ítems	95
B3	Informe de Validación	96
B4	Consultas Overpass para el Corregimiento Omar Torrijos	97
B5	Representación Gráfica del Contenido del Archivo de Omar Torrijos	97
B6	Parámetros de Generación de Imágenes	98
B7	Ejemplo de Etiquetado	98
B8	Parámetros Augmentación de Imágenes con Python	99
B9	Prompt en Memex	99
B10	Panel de Nuevas Inferencias	100
B11	Carga de Imágenes Desde la Computadora	100
B12	Descarga de Imágenes Desde Google Drive	101
B13	Vista Previa de la Predicción	101
B14	Chat Copiloto GPT-4o Generando Respuesta Textual y Visual	102
B15	Gráfico de Distribución de Residuos	102
B16	Matriz de Confusión y Métricas Relativas	103
B17	Prompts de Creación de Métricas de Fidelidad	103
B18	Fórmula de Precisión para Modelos de Inteligencia Artificial	104
B19	Fórmula de Recall para Modelos de Inteligencia Artificial	104
B20	Formula de Mean Average Precision (MAP)	104
B21	Formula de Matriz de Confusión	105
B22	Fórmula de Precisión General	105
B23	Fórmula de Execution Accuracy (EX)	105
B24	Fórmula de Visual-Textual Similarity (VTS)	106

B25	Matriz de Confusión y Métricas Estándar	63
B26	Resultados Mean Average Precision (MAP) Durante los Entrenamientos	64
B27	Distribución Porcentual de Acumulaciones de Residuos	65
B28	Cantidad de Acumulaciones Detectadas por Corregimientos	65
B29	Matriz de Confusión Imágenes Enero 2024	66
B30	Resultados de Métricas del Modelo	66
B31	Resultados de Execution Accuracy (EX)	67
B32	Resultados de Visual-Textual Similarity (VTS)	68



REPÚBLICA DE PANAMÁ

UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA

**PROTOTIPO CON VISIÓN COMPUTACIONAL EXPLICABLE-SEMÁNTICO  
PARA DETECCIÓN DE ACUMULACIÓN DE RESIDUOS EN LAS CALLES DE  
SAN MIGUELITO, CIUDAD DE PANAMÁ**

Autor: Jorge Andres Gregory Guette

Tutor: José Rivera Castro

Año: 2024

**RESUMEN**

El objetivo general de este estudio es implementar un prototipo con visión computacional explicable-semántico para la detección de acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá, con la finalidad de aportar una herramienta innovadora que facilite la planificación y gestión de recolección de residuos. Esta investigación se desarrolló como tipo aplicada, con enfoque cuantitativo y diseño cuasiexperimental simulado. La población utilizada son todas las imágenes satelitales del distrito de San Miguelito, Ciudad de Panamá, y como muestra se seleccionaron imágenes satelitales de San Miguelito de los períodos de abril de 2023 para entrenamiento y enero de 2024 para validación general. En la investigación, los instrumentos fueron: ficha de observación para el conteo manual, y un informe de evaluación con: precisión general del modelo, Execution Accuracy (EX) y Visual-Textual Similarity (VTS). En cuanto al resultado, el prototipo alcanzó un rendimiento general de 86% en detección, con el conteo sistemático de 322 detecciones, señalando a los corregimientos como Belisario Frías con 75 y Omar Torrijos con 67, donde se detectaron más cúmulos de residuos. En la fidelidad explicativa, con EX de 88%, mientras que con VTS alcanzó un promedio de puntos de 23.0 superando los umbrales propuestos en la investigación, por lo tanto, se evidenció la factibilidad técnica del modelo para el contexto urbano de San Miguelito. En conclusión, el estudio confirma la hipótesis de que un prototipo de visión computacional explicable-semántico permite identificar de forma fiable acumulaciones de residuos en las calles de San Miguelito, Ciudad de Panamá.

**Palabras clave:** Visión Computacional, Inteligencia Artificial Explicable, Residuos Sólidos Urbanos, San Miguelito, Imágenes Satelitales.



**REPUBLIC OF PANAMA**  
**INTERNATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY**  
**FACULTY OF COMPUTER SCIENCE AND TECHNOLOGY**

**PROTOTYPE WITH EXPLAINABLE-SEMANTIC COMPUTER VISION FOR THE  
DETECTION OF WASTE ACCUMULATION IN THE STREETS OF SAN MIGUELITO,  
PANAMA CITY**

Author: Jorge Andres Gregory Guette

Tutor: José Rivera Castro

Year: 2024

**ABSTRACT**

The general objective of this study is to implement an explainable-semantic computer vision prototype for detecting waste accumulation on the streets of San Miguelito, Panama City, with the purpose of providing an innovative tool that facilitates waste collection planning and management. This research was conducted as an applied type, with a quantitative approach and a simulated quasi-experimental design. The population used includes all satellite images of the San Miguelito district, Panama City, and as a sample, satellite images of San Miguelito from the periods of April 2023 for training and January 2024 for general validation were selected. In the research, the instruments were: an observation sheet for manual counting, and an evaluation report with: general model accuracy, Execution Accuracy (EX), and Visual-Textual Similarity (VTS). Regarding the result, the prototype achieved an overall performance of 86% in detection, with the systematic count of 322 detections, identifying the townships of Belisario Frías with 75 and Omar Torrijos with 67 as those with the highest waste accumulations. In explanatory fidelity, it obtained an EX of 88%, while with VTS it reached an average score of 23.0, exceeding the thresholds proposed in the research. Therefore, the technical feasibility of the model for the urban context of San Miguelito was demonstrated. In conclusion, the study confirms the hypothesis that an explainable-semantic computer vision prototype allows the reliable identification of waste accumulations on the streets of San Miguelito, Panama City.

**Keywords:** Computer Vision, Explainable Artificial Intelligence, Urban Solid Waste, San Miguelito, Satellite Images.

## INTRODUCCIÓN

El distrito de San Miguelito, Ciudad de Panamá, enfrenta una alerta sanitaria emitida en enero de 2024 (Ministerio de Salud de la República de Panamá [MINSA], 2024) debido a la gran acumulación de residuos en las calles, distribuidos por los 9 corregimientos que lo conforman. Este problema en el distrito agrava el ya existente en Panamá, donde el 12% de los residuos sólidos municipales del país no son recogidos por los servicios de aseo, señalando que estarían apilados en las calles o en vertederos clandestinos (Alarcón Montero et al., 2023). Este trabajo propone un prototipo de visión computacional explicable-semántico para detectar esas acumulaciones en las calles de San Miguelito, teniendo como guía de esta investigación la pregunta: ¿En qué medida puede implementarse un prototipo de visión computacional explicable-semántico que permita detectar la acumulación de residuos en San Miguelito? Además, plantea el objetivo general de implementar el prototipo, desprendiendo tareas concretas: construir el sistema y contabilizar la acumulación en imágenes; medir el rendimiento con métricas estándar; y evaluar la fidelidad explicativa-semántica.

La hipótesis de esta investigación sostiene que el prototipo permitirá identificar de forma fiable la acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá; teniendo como alcance del estudio el diseño, la implementación y la evaluación del prototipo en el contexto de San Miguelito, Ciudad de Panamá, y no incluye el despliegue en campo. En cuanto al tipo de investigación realizada, según su finalidad es aplicada; y según el enfoque, será investigación cuantitativa, con un diseño cuasiexperimental simulado. Este estudio se llevará a cabo cada parte en las consecuentes secciones.

## **PLANTEAMIENTO DEL PROBLEMA**

En el mundo entero, se señala una problemática alarmante que es la gran acumulación de residuos en las calles. El 70% de los desechos termina en basurales y rellenos sanitarios y solo el 19% se recicla (Greenpeace Colombia, 2023). Para América Latina y el Caribe, se presenta una situación similar en la acumulación de residuos que terminarán en las calles, con 145,000 toneladas por día (Programa de las Naciones Unidas para el Medio Ambiente, 2020). Por lo tanto, esta situación ha generado iniciativas de prototipos con inteligencia artificial para afrontar dicho problema, monitoreando basurales a cielo abierto (Fundación Bunge y Born, 2021). En Panamá, se presenta que el 12% de los residuos sólidos municipales del país no son recogidos por los servicios de aseo, señalando que estarían apilados en las calles o vertederos clandestinos (Alarcón Montero et al., 2023); debido a eso han surgido proyectos de inteligencia artificial para identificar residuos como la empresa REDOXIGEN, Inc., que gana el Premio Nacional a la Innovación Empresarial 2020, con un sistema con visión computacional para reconocer basura extraída de los ríos (Secretaría Nacional de Ciencia, Tecnología e Innovación [SENACYT], 2020).

En la Región de Salud del Distrito de San Miguelito, Ciudad de Panamá, en enero de 2024, se declaró alerta sanitaria ante el problema de acumulación masiva de residuos en las calles, dentro del comunicado la Autoridad señala que la causa principal del problema es la ineficiencia en el proceso de recolección de basura por parte del operador del servicio (MINSA, 2024). De persistir la situación, aumentarán drásticamente los ya existentes casos de dengue en el Distrito de San Miguelito y toda la Ciudad de Panamá (MINSA, 2025). Este estudio propone detección automática con imágenes satelitales de acumulaciones de residuos en las calles de San Miguelito mediante la implementación de un prototipo con visión computacional explicable-semántico, además, esta investigación servirá como punto de partida para mejorar la gestión en el proceso de recolección de residuos.

## **Justificación**

El problema de acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá, ya es una realidad debido a la alerta emitida por las autoridades de salud pública en 2024, causando daños críticos a nivel ambiental y de salud (MINSA, 2024). Este estudio, como valor teórico, propone la implementación de un prototipo web basado en visión computacional explicable-semántico, con la capacidad de visualizar, detectar y explicar con lenguaje natural por qué decidió identificar el punto como cúmulo de residuos de forma automática. A nivel práctico, esta herramienta ayuda a facilitar la toma de decisiones de las autoridades en la planificación del cronograma de rutas, priorizando las zonas más críticas; igualmente, tendría beneficios económicos al reducir costos operativos.

En cuanto al valor social, contribuye a la mejora en las condiciones de vida de los sectores más vulnerables del distrito de San Miguelito, Ciudad de Panamá; mientras que, en lo metodológico, este sistema introduce explicabilidad en la automatización de la detección al describir en lenguaje natural los factores que influyeron en la predicción, usando un Large Language Model (LLM) y los resultados de los objetos detectados en la imagen, lo cual fortalece la validación correcta en el prototipo. Por las razones anteriores, el proyecto es plenamente justificable como pertinente, innovador y replicable en contextos urbanos.

## **Pregunta de investigación**

¿En qué medida puede implementarse un prototipo de visión computacional explicable-semántico que permita detectar la acumulación de residuos en San Miguelito, Ciudad de Panamá?

## **Objetivos**

### **Objetivo General**

Implementar el prototipo con visión computacional explicable-semántico para detección de acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá.

### **Objetivos Específicos**

En este estudio se identifican los siguientes objetivos específicos:

- Construir el prototipo con visión computacional explicable-semántico: fase de diseño y entrenamiento.
- Contabilizar la acumulación de residuos en las imágenes satelitales de San Miguelito, Ciudad de Panamá: fase de diseño y entrenamiento.
- Medir el rendimiento general del modelo de visión computacional explicable-semántico: fase de evaluación de generalización.
- Evaluar la fidelidad explicativa-semántica del prototipo con visión computacional: fase de evaluación simulada.

## **Hipótesis**

El prototipo de visión computacional explicable-semántico permitirá detectar de forma fiable la acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá.

## **Alcance y Delimitación del Proyecto**

El presente estudio tiene como alcance el diseño, implementación y evaluación de un prototipo web con visión computacional explicable-semántico para la detección de acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá. Dicho prototipo se limita a procesar y analizar imágenes satelitales obtenidas de la aplicación Google Earth Pro, demarcadas por los límites administrativos obtenidos de la plataforma Overpass Turbo, y a generar explicaciones en lenguaje natural que permiten entender por qué un área específica fue clasificada como acumulación de residuos. El estudio no contempla la implementación en campo del prototipo en la gestión operativa municipal de residuos, ni la integración con alguna interfaz de aplicaciones de la plataforma del operador responsable de la recolección de residuos.

Tampoco incluye la verificación en campo de las detecciones realizadas, por lo que los resultados están restringidos a la validación con imágenes satelitales segmentadas en teselas de 150 metros con un zoom de 213 dentro de Google Earth Pro. Geográficamente, la investigación se delimita al distrito de San Miguelito y a sus nueve corregimientos urbanos, áreas rurales y boscosas; en cuanto a la temporalidad, las imágenes usadas serán de abril de 2023 para entrenamiento y enero de 2024 para validación, debido a que la herramienta Google Earth Pro solo tiene imágenes disponibles en esos últimos dos períodos.

En el aspecto técnico, el alcance se restringe a la detección de acumulaciones de residuos en espacio cielo abierto, como calles, aceras y lotes baldíos, sin abordar la clasificación de tipos de desechos ni su cuantificación volumétrica. De esta manera, los resultados de este proyecto deben interpretarse como una prueba de concepto que demuestra la factibilidad técnica al combinar visión computacional y explicabilidad semántica en la gestión de residuos en contextos urbanos. En futuras investigaciones se podrá ampliar este alcance mediante la incorporación de más zonas geográficas, una temporalidad extendida y la realización de pruebas reales.

## **MARCO TEÓRICO**

El marco teórico es el componente que proporciona contexto conceptual y teórico para el estudio a realizar dentro de las investigaciones ya elaboradas, creando conexiones o resaltando las brechas que el estudio pretende llenar. También indica que su estructura general es la siguiente: antecedentes, bases teóricas y términos básicos (Tarrillo et al., 2024). A continuación, esta investigación continúa con esa organización, iniciando con investigaciones previas basadas en las variables de investigación y subdivisiones como la acumulación de residuos, visión computacional y visión computacional explicable.

La base teórica contiene cinco conceptos principales: visión computacional, inteligencia artificial explicable, modelos de lenguaje de gran escala, Text-to-SQL basado en modelos de lenguaje de gran escala y las fases en proyectos de inteligencia artificial explicable, además de términos básicos importantes para la investigación. Cada antecedente y teoría expuesta en esta sección tendrá su párrafo explicando el aporte que tendrá en esta investigación.

### **Antecedentes**

#### **Acumulación de residuos en las calles**

Isaza et al. (2025) realizaron una investigación titulada “Análisis del Área y Calidad del Suelo del Relleno Sanitario del Jobo de Antón en la Provincia de Coclé Mediante el uso de la Teledetección” en Panamá. El objetivo general de la publicación es utilizar la Teledetección como herramienta para describir el aumento de área por año en la recolección de los desechos sólidos del suelo en el relleno sanitario del Jobo de Antón en la provincia de Coclé; como muestra se utilizaron cuatro imágenes correspondientes a los años 2014, 2020 y 2023; el diseño de la investigación es el estudio descriptivo longitudinal basado en Teledetección multitemporal de los años 2014, 2018, 2020 y 2023; en cuanto a instrumentos, se

usó Google Earth Pro, Copernicus Data Space EcoSystem y el Software ArcGIS Pro.

Los resultados fueron la impactante degradación del suelo a pesar de la disminución del área de ocupación en el año 2023 con 0.77 hectáreas; el estudio de Isaza et al. (2025) utiliza la función historial de Google Earth Pro para retroceder a fechas puntuales con disponibilidad de imágenes satelitales en esa zona como conjuntos de datos para el análisis descriptivo por parte del investigador, en esta investigación se usará esta opción para obtener el conjunto de imágenes de entrenamiento disponible del mes de abril de 2023 y como conjunto de valores reales, enero de 2024.

Bernal (2024) realizó un estudio titulado: “Situación del Manejo de Residuos Municipales en el Área Metropolitana de la Ciudad de Panamá y su Influencia en el Desarrollo de Vectores”. El objetivo general del estudio fue describir la relación entre el manejo de residuos y la existencia o no de vectores infecciosos; la muestra empleada en el estudio son estaciones de contenedores colocadas por las autoridades tipo C4 a C1 y también hizo una clasificación de los residuos de mayor avistamiento, como orgánicos, botellas plásticas y cartón; su diseño investigativo fue experimental, cuantitativo y con el método Estudio, Diseño, Ejecución, Revisión (EDER); los instrumentos utilizados en el estudio fueron un formulario estructurado de recolección de datos, que incluía las generales de los sitios, caracterización de los contenedores, caracterización de residuos en presencia de vectores y otros dos formularios: solicitud de código de investigación responsables (COIR) y condiciones de prevención de riesgos laborales; además, se utilizaron dos aplicaciones: MyGPS Coordinates para etiquetar con georreferencias las estaciones de la muestra y el InfoStatv2023 para análisis estadístico.

Los resultados obtenidos del estudio mostraron que las especies de mosquitos con mayor número de presencia en los vectores de la muestra fueron *Aedes aegypti* (Dengue), *Aedes albopictus* (Chikungunya) y *Culex quinquefasciatus* (Mosquito común), demostrando la relación epidemiológica de los

vectores en el Área Metropolitana de la Ciudad de Panamá y el manejo de residuos. El estudio de Bernal (2024) expone la relación existente entre los vectores de contaminación y su relación con la gestión de residuos; esta investigación utiliza esa correlación ya demostrada al medir las variables dependientes e independientes de forma cuantitativa.

Reche et al. (2022) realizaron una investigación titulada “Vehicular Traffic in Urban Areas: Health Burden and Influence of Sustainable Urban Planning and Mobility” en 12 ciudades de 9 países europeos. El objetivo general era determinar qué características del entorno urbano influyen en la contaminación por tráfico y evaluar los beneficios de salud al tener intervenciones de mitigación; la muestra consistió en indicadores urbanos como el tráfico diario, el número de paradas de transporte, el porcentaje de áreas verdes y los kilómetros de ciclovías recolectados para cada ciudad, que se descargaron vía Overpass Turbo; además, el estudio es comparativo con una evaluación multicriterio de indicadores urbanos; los instrumentos usados son el software Geographic Information System (GIS) para calcular los indicadores y el Environmental Benefits Mapping and Analysis Program – Community Edition (BenMAP-CE) para computar Health Impact Assessment (HIA).

En conclusión, los resultados obtenidos en la investigación fueron que se identificaron dos tendencias principales: las ciudades con peor distribución de paradas de transporte tienen menos kilómetros de carriles bici, sugiriendo baja accesibilidad ciclística, y el mayor porcentaje de áreas verdes se asoció a más viajes en bicicleta. Esta investigación utilizará el mismo método de obtención de datos urbanos vía Overpass Turbo, descargando los límites administrativos de cada corregimiento del distrito de San Miguelito, Ciudad de Panamá.

Santos y Dorneles (2022) realizaron una investigación titulada “Enriquecer Bases de Datos con Información Ambiental Georreferenciada Mediante la Extracción de Datos Procedentes de Archivos KML Disponibles en la Web” en Brasil. El objetivo general consistía en enriquecer bases de datos con información

ambiental georreferenciada mediante la extracción de datos procedentes de archivos de Keyhole Markup Language (KML) disponibles en la web; la muestra fue obtenida de datos geográficos en formato KML con polígonos representando áreas de interés de los repositorios en línea ([iaria.org](http://iaria.org)), además usaron datos públicos de Brasil con centros de distritos y condados para enriquecer información ambiental; respectivamente, el diseño del estudio fue experimental aplicativo con el desarrollo de un extractor web que descarga y procesa archivos KML que se integran en bases de datos ambientales públicas para agregar información geográfica.

Finalmente, los resultados obtenidos, se mostró que su ejemplo práctico generó datos enriquecidos que incluyen coordenadas geográficas que servirían para apoyar inspecciones en áreas rurales en Brasil. La contribución de la investigación de Santos y Dorneles (2022) es la utilización de los archivos KML para representar áreas de interés y poder obtener datos para georreferenciación; en el prototipo se usarán estos archivos para delimitar por áreas en Google Earth Pro de 150 metros con un zoom de 213, basadas en los límites administrativos vía Overpass Turbo para generar capturas de imágenes que servirán como insumo para la muestra.

Pitty y Migdaris (2021) confeccionaron un estudio titulado: "Evaluación de Impactos y Problemáticas Existentes en las Comunidades Aledañas al Relleno Sanitario Cerro Patacón, Ciudad de Panamá, Panamá". El objetivo general de la investigación fue evaluar los impactos que se generan en las inmediaciones del vertedero Cerro Patacón, producto de la gestión y disposición de los desechos sólidos en ese lugar; la muestra consistió en 45 encuestas a la comunidad de Kuna Nega, seleccionada por su cercanía y su mayor afectación por el relleno; además, el tipo de estudio realizado es descriptivo, experimental y cuantitativo, dividido en cinco fases: la primera es una investigación de la situación actual en el manejo de residuos, la segunda fue la clasificación de tipos de residuos y la última es la identificación de vectores infecciosos como Dengue, Fiebre Amarilla y Zika; los

instrumentos utilizados fueron un formulario con datos generales del sitio, caracterización de residuos y encuestas a la comunidad.

En cuanto a los resultados obtenidos por el estudio, el impacto más sentido por los encuestados es la contaminación del aire, con el 89%, y la contaminación de cuerpos de agua. Los resultados de Pitty y Migdaris (2021) demuestran el alcance de la contaminación del aire causada por el vertedero Cerro Patacón, que llega a comunidades aledañas como Kuna Nega, que se encuentra en las faldas del cerro. En esta investigación se tomarán en cuenta las zonas cercanas a las acumulaciones de residuos, ya que su identificación por el prototipo será por zonas de 150 metros y con un solapamiento entre segmentos del 12% en una imagen satelital.

## **Visión Computacional**

Miller (2025) efectuó una investigación titulada “Exploring the Potential of No-Code Artificial Intelligence in Identifying Periapical Radiolucent Lesions” en Estados Unidos. El objetivo principal fue emplear una plataforma no-code para desarrollar un modelo de inteligencia artificial diseñado para identificar lesiones radiolúcidas periapicales (mancha oscura alrededor de la raíz de un diente) en radiografías periapicales; la muestra consistió en 312 radiografías periapicales retrospectivas y reportes radiológicos obtenidos dentro de los seis meses previos. De esas imágenes, 180 tenían al menos una lesión periapical radiolúcida y 132 no tenían lesiones; el diseño de investigación fue retrospectivo con desarrollo de modelo de inteligencia artificial; en los instrumentos, utilizó la plataforma LandingLens no-code, basada en la nube para entrenar el modelo, etiquetar imágenes y hacer inferencias, además de métricas diagnósticas como sensibilidad, especificidad, precisión, exactitud y F1 score.

En conclusión, los resultados determinaron que el modelo de inteligencia artificial sin código puede identificar con buena precisión las radiolucencias periapicales en radiografías periapicales. El uso de la herramienta LandingLens

por el autor Miller (2025) será empleado en esta investigación para etiquetar las imágenes y entrenar el modelo de visión computacional que se implementará en el prototipo para realizar inferencias sobre la muestra de imágenes satelitales con acumulaciones de residuos en el distrito de San Miguelito, Ciudad de Panamá.

Cárdenas et al. (2024) confeccionaron un estudio titulado: “Modelo de un Sistema de Alerta Basado en Visión Artificial para la Prevención de Aglomeraciones en el Transporte Público” en Panamá. El objetivo principal de la investigación era desarrollar un modelo para la detección en tiempo real de aglomeraciones y proporcionar a los operadores de transporte información sobre la cantidad de usuarios en los buses; el diseño es experimental-aplicativo, ya que se manipularon condiciones y pruebas con diferentes escenarios para validar el prototipo; la población y muestra están compuestas por tres pruebas: 12 personas enfocadas con una cámara colocada de costado, 5 personas enfocadas desde diferentes ángulos y con variación de iluminación, y 14 personas en un pasillo estrecho enfocadas con la cámara en la parte superior de la entrada, para un total de 31 personas; los instrumentos utilizados fueron cámara web, librerías de visión artificial, deep learning y pruebas experimentales controladas.

Los resultados obtenidos en la investigación el sistema propuesto detectó con una precisión global del 84.04%, demostrando la posibilidad de optimizar el flujo de pasajeros. El estudio realizado por Cárdenas et al. (2024) tomó como métrica principal la precisión global del modelo, con un 84.04%, para demostrar que su prototipo tenía la posibilidad de optimizar el flujo de pasajeros; por lo tanto, en esta investigación se tomará como base la precisión global del modelo para medir el rendimiento al detectar acumulación de residuos en la fase de entrenamiento y validación con un 84.04% de umbral de aceptación.

Santamaría et al. (2022) efectuaron una investigación titulada: “Reconocimiento de Tipos de Serpiente en Panamá por Medio de Visión Artificial”. El objetivo general fue desarrollar un sistema de visión artificial utilizando redes neuronales; la población y muestra fueron un mínimo de 100 fotos de las distintas

especies venenosas y no venenosas de serpientes en Panamá; el tipo de investigación fue diseño experimental-aplicado porque se creó y probó una herramienta tecnológica en condiciones controladas; como instrumento utilizado, se empleó el aplicativo móvil de visión artificial con el framework TensorFlow como medio de recolección, procesamiento y clasificación de serpientes venenosas o no.

Por último, los resultados de la investigación fueron que el uso de la visión artificial con redes neuronales detectó con precisión las imágenes nuevas de serpientes, clasificándolas como venenosas o no; el estudio de Santamaría et al. (2022) muestra el uso del framework de código abierto TensorFlow en la aplicación móvil con visión artificial. Este estudio usará de forma similar un framework de código abierto que es MMDetection empleado por la plataforma en línea LandingLens.

### **Visión Computacional para Detectar Residuos**

Cañete (2024) confeccionó un estudio titulado: “Detección de Basurales a Cielo Abierto en Nuestra Región Mediante el Uso de Inteligencia Artificial: Generación de los Datos Locales” para las ciudades de Paraná, Santa Fe y Rosario en Argentina. Los objetivos de la investigación son: generar un conjunto de imágenes satelitales locales con basurales señalados, determinar si es viable el uso de Convolutional Neural Network (CNN) para detectar basurales a cielo abierto, y comprobar el rendimiento de las CNN en la detección de los basurales a cielo abierto; la muestra fueron imágenes de ciudades como Rosario con 900 imágenes, Santa Fe con 600 imágenes y Paraná con 600 imágenes; el diseño de la investigación es cuantitativo por el manejo del porcentaje de basurales por ciudad y experimental, ya que introduce mediciones comparativas del rendimiento de los modelos de inteligencia artificial en diferentes condiciones; en cuanto a instrumentos, utilizó un sistema GIS de código abierto llamado Quantum Geographic Information System (QGIS) para descargar imágenes satelitales y el

uso de hojas de cálculo de Google for Education para el etiquetado manual de 1 con basurales y cero sin basurales.

Para finalizar el resultado del estudio fue de la ciudad de Rosario, con 600 imágenes, se detectaron 39 basurales; de Santa Fe, con 600 imágenes, se detectaron 8 basurales; y de Paraná, con 600 imágenes, se detectaron 39 basurales, determinando una mayor concentración de residuos en Paraná, posiblemente asociada a su geografía y situación socioeconómica. Esta investigación tomará como origen la obtención de imágenes satelitales por medio del sistema GIS de código abierto y gratuito como Google Earth Pro, al igual que el estudio de Cañete (2024), que utilizó como origen de las imágenes satelitales la herramienta QGIS. También, Cañete (2024) utilizó hojas de cálculo de Google for Education para etiquetar manualmente las imágenes con observación sistemática de cada imagen, igualmente, en este estudio se usará una ficha de observación sistemática almacenando en Excel las etiquetas manuales de las imágenes satelitales.

Torres y Fraternali (2023) efectuaron una investigación titulada: “Aerial Waste Dataset for Landfill Discovery in Aerial and Satellite Images” en Lombardía. El objetivo general es desarrollar y poner a disposición pública un dataset de alta calidad para facilitar el entrenamiento, validación y pruebas de modelos de Deep Learning aplicados al monitoreo ambiental; la muestra utilizada fue de 10,434 imágenes aéreas y satelitales (3,478 positivas y 6,956 negativas) obtenidas de Agenzia per le Erogazioni in Agricoltura (AGEA), WorldView-3 y Google Earth; el diseño de investigación fue desarrollo experimental y validación técnica de un dataset para clasificaciones binarias y multiclase; los instrumentos utilizados en la herramienta fueron annotation ODIN, clasificador (Residual Network with 50 layers) RESNET50 con Feature Pyramid Network para validación técnica.

En definitiva, los resultados, la investigación arrojó que el clasificador binario RESNET50 entrenado con todas las fuentes obtuvo 87.99% de precisión promedio, 80.70% en F1-score, 81.89% de precisión y 79.54% de sensibilidad (recall); el

modelo entrenado solo con ortofotos AGEA: 94.5% de precisión promedio, 88.2% en F1-score. Igualmente, se usó el modelo para asistir a analistas ambientales identificando 69 sitios sospechosos, confirmando su aplicabilidad práctica. Este estudio tomará una muestra pequeña de imágenes de residuos del conjunto de datos público AerialWaste usadas en la investigación de Torres y Fraternali (2023) para ampliar el conjunto de imágenes positivas y aumentar la precisión general del modelo.

Pascasio et al. (2022) hicieron un estudio titulado: “Implementación de Redes Neuronales para la Clasificación de Desechos Dentro de un Cesto Inteligente” en Panamá. El objetivo principal es la implementación de un dispositivo para la clasificación automática de desechos dentro de un cesto inteligente; el diseño de la investigación fue aplicativo con enfoque experimental y tecnológico, ya que implementa y valida un prototipo de cesto inteligente; la población y muestra fue de 2,527 imágenes de un conjunto de ejemplos ya descargado, y los autores añadieron 380 imágenes adicionales para equilibrar las clases, con lo cual la muestra total llegó a 2,907 imágenes; los instrumentos fueron en hardware: Raspberry Pi 4 B+, cámara, sensores magnéticos, motores, controladores y Arduino Mega; de software se usó TensorFlow y Keras, además de modelos probados como: Mobile Neural Network (MOBILENET), Visual Geometry Group 16-layer Network (VGG16), RESNET50, Inception Residual Network Version 2 (INCEPTIONRESNETV2); también, emplearon Android Studio y Firebase Realtime Database.

Finalmente, los resultados de la investigación fue que la comparación de las métricas de precisión para cada modelo y fueron: MOBILENET: precisión 64 %; VGG16: precisión 88 %; RESNET50: precisión 91 %; INCEPTIONRESNETV2: precisión 80 %. Los autores seleccionaron a RESNET50 para implementarla en el prototipo por su mayor precisión. El prototipo implementado por Pascasio et al. (2022) utiliza una base de datos en tiempo real como Firebase Realtime Database para sincronizar las predicciones de la aplicación móvil con el backend, el prototipo de esta investigación no requiere sincronización en tiempo real; por lo tanto, se

utilizará una base de tipo Structured Query Language (SQL) ligera de código abierto, como SQLite, para almacenar las predicciones del modelo y las ubicaciones geoespaciales.

### **Visión Computacional Explicable**

Li et al. (2024) realizaron una investigación titulada “Explainable Visual Question Answering with Multimodal Rationales” en los Países Bajos. El objetivo principal del estudio es proponer un nuevo modelo denominado Multimodal Rationales for Visual Question Answering (MRVQA) que genera respuestas con justificación textual y visual para mejorar la explicabilidad y la confianza; la muestra utilizada fue de 33,726 pares pregunta-respuesta derivados de 20,367 imágenes del dataset Visual Question Answering Explanation Synthetic (VQA-E-SYN); el diseño del estudio ha sido experimental comparativo con métodos previos; en cuanto a los instrumentos utilizados, es el modelo MRVQA basado en Transformer, Generative Pre-trained Transformer 2 (GPT-2) para generar explicaciones textuales y explicaciones visuales con Grounded Denoising Object Detection Transformer (GROUNDINGDINO).

Para concluir, los resultados logrados en la investigación de Li et al. (2024) del modelo alcanzaron una exactitud del 75.01% en respuestas en VQA-E-SYN, y además superaron métodos anteriores en métricas de explicación como Bilingual Evaluation Understudy – 4-gram (BLEU-4), Consensus-based Image Description Evaluation (CIDEr), Recall-Oriented Understudy for Gisting Evaluation (ROUGE), Semantic Propositional Image Caption Evaluation (SPiCE) y Visual–Textual Similarity (VTS). En la investigación de Li et al. (2024) utilizaron como medida base comparativa de la propuesta de fidelidad visual y textual en las respuestas la métrica VTS con el promedio de las últimas pruebas como umbral de aceptación, ya que VTS busca combinar la similitud textual y visual indicando que no solo basta que el texto suene bien, sino que sea respaldado visualmente; en este estudio se medirá la fidelidad de las explicaciones aplicando la métrica VTS mediante pruebas

automatizadas dentro de la aplicación, generando medias VTS que exponen los resultados de las evaluaciones.

Zhou et al. (2024) ejecutaron una investigación titulada “Image-of-Thought Prompting for Visual Reasoning Refinement in Multimodal Large Language Models” en China. El objetivo general fue proponer y evaluar el método Image-Of-Thought (IOT) para mejorar la explicabilidad visual en modelos multimodales de lenguaje, integrando razonamiento visual y textual; la muestra manejada en el estudio consistía en conjuntos de datos de Visual Question Answering (VQA); el diseño fue un estudio experimental comparativo, basado en pruebas con modelos de referencia como GPT-4 o Gemini-Pro; los instrumentos empleados son: Fast Segment Anything Mode (FASTSAM) para segmentación y GROUNDINGDINO para detección, junto a la Application Programming Interface (API) de Generative Pre-trained Transformer 4 (GPT-4) con el parámetro de temperatura igual a cero, simulando razonamiento determinista.

Finalmente, los resultados determinaron que IOT prompting mejoró el rendimiento en precisión del razonamiento multimodal comparado con los métodos tradicionales; además, el método redujo alucinaciones y fortaleció la interpretabilidad. El estudio de Zhou et al. (2024) utilizó como instrumento la API GPT-4 con el parámetro de temperatura igual a cero, que vuelve al modelo su razonamiento más determinista, sin embargo, disminuyendo la creatividad; esta investigación realizará la misma parametrización del módulo de explicabilidad semántica con la API de Generative Pre-trained Transformer 4 omni (GPT-4o) y el parámetro de temperatura con valores cercanos a cero.

Lu et al. (2022) desarrollaron un estudio titulado “Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering” en Estados Unidos. El objetivo general de la investigación fue presentar Science Question Answering (SCIENCEQA), un nuevo benchmark con preguntas para diseñar modelos de lenguaje que generen explicaciones tipo Chain of Thought (COT), buscando mejorar el razonamiento e interpretabilidad en ciencias; en

cuanto a la muestra utilizada, son 21,208 ejemplos de preguntas multimodales de ciencias conformadas por texto, imágenes y diagramas de nivel primaria o secundaria; el diseño de la investigación fue un estudio experimental comparativo con modelos base VQA, UnifiedQA (Unified Question Answering) y Generative Pre-trained Transformer 3 (GPT-3), midiendo la precisión de respuesta y calidad de las explicaciones.

En definitiva, los resultados de la investigación arrojaron que las cadenas de pensamiento mejoran el rendimiento, permitiendo aprender con un 40% menos de datos y generando explicaciones de los casos que cumplen con el estándar humano en un 65.2%. En la investigación de Lu et al. (2022), el razonamiento encadenado es para texto explicativo del examen y el aporte en el prototipo de esta investigación, el razonamiento encadenado es basado en la estructuración de consulta a la base de datos con GPT-4o, ejecución de sentencias, ventana de contexto y explicación narrativa.

Selvaraju et al. (2017) efectuaron una investigación titulada “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization” en Estados Unidos. El objetivo general fue proponer a GRAD-CAM (Gradient-weighted Class Activation Mapping) como técnica para generar explicaciones de modelos de redes CNN sin modificar su estructura, añadiendo transparencia en las decisiones en la clasificación de imágenes; la muestra del estudio eran datasets estándar como ImageNet (ILSVRC-15), PASCAL VOC 2007, subconjuntos de VQA y Dense Captioning (DENSECAP); el diseño de la investigación es experimental comparativa de GRAD-CAM con otros métodos como Class Activation Mapping (CAM), Backpropagation, Deconvolution, estudios de usuarios y Compositional Multi-modal Weakly-supervised Pretraining (C-MWP); los instrumentos fueron Convolutional Neural Network (CNN) como VGG-16, Named after Alex Krizhevsky (ALEXNET), Residual Network (RESNET), Neural Talk version 2(NEURALTALK2), VQA y GRAD-CAM, además de Amazon Mechanical Turk para estudios de percepción y confianza.

Al final, los resultados de la investigación mostraron que GRAD-CAM localiza mejor las zonas importantes en imágenes que otros métodos, que los usuarios confiaron más en los modelos cuando se mostraban explicaciones de GRAD-CAM, y que en VQA, mostró explicaciones más claras y cercanas a la atención humana. Este estudio se agregó dentro de los antecedentes por la contribución significativa en el campo de Inteligencia Artificial Explicable al generalizar GRAD-CAM como método post-hoc aplicable para cualquier red neuronal convolucional y no solo redes que tienen scope global, sirviendo como base replicable a investigaciones posteriores de explicabilidad en inteligencia artificial.

En el Anexo A, Tabla A1 se muestra el resumen de los antecedentes seleccionados, organizados con base en el año, autor, título de la investigación y el aporte que cada estudio ofrece a la presente investigación.

## **Bases Teóricas**

### **Visión Computacional**

El concepto de visión computacional, según los autores Bi et al. (2022), es un campo de la inteligencia artificial que permite extraer, analizar y comprender el contenido en imágenes. También, Moin (2023) indica que es una disciplina que permite a las máquinas interpretar y analizar información visual por medio de algoritmos que simulan la vista humana. Por último, Karn (2021) la define como un área de investigación de la inteligencia artificial que ayuda a ver a las computadoras infiriendo información abstracta del mundo a partir de imágenes. En este estudio, se apoya en la definición de Bi et al. (2022) por tener las principales funciones de la visión computacional, que son las de extraer, analizar y comprender el contenido visual.

## **Tipos de Visión Computacional**

Según Navarrete (2024), los tipos de visión computacional son: clasificación de imágenes, que permite una o varias imágenes extraer una clase por imagen; la detección de objetos, que muestra cada una de las clases y arroja las coordenadas de cada una para poder localizarlas en la imagen; la segmentación, que va a encontrar distintos elementos en la imagen, indicando el grupo de píxeles que representan cada objeto. Además, Sarkar et al. (2025) define los tipos de visión computacional en: razonamiento relacional, entender relaciones entre entidades espaciales; razonamiento simbólico: manipulación de representaciones simbólicas derivadas de grafos, programas o lógica; razonamiento temporal: inferencias a través del tiempo con secuencias o evolución de estados; razonamiento causal: identificar causas y efectos; razonamiento de sentido común: usar conocimiento general e inferir motivaciones no visibles directamente.

Esta investigación selecciona el tipo de visión computacional detección de objetos en la imagen definido por el autor Navarrete (2024), al describir exactamente lo que va a realizar el prototipo con visión computacional, que es la detección de acumulación de residuos en la imagen satelital, resaltando lo detectado y guardando la ubicación en píxeles.

## **Inteligencia Artificial Explicable**

### **Definición**

Según Barredo et al. (2020), la Explainable Artificial Intelligence (XAI) es un campo emergente que propone agregar transparencia y comprensibilidad a modelos de aprendizaje automático complejos, llamados modelos de “caja negra”, para convertirlos en sistemas capaces de comunicar cómo y por qué se produce una decisión. También, Palacio et al. (2021) lo definen como un campo interdisciplinario que transforma modelos opacos en sistemas híbridos explicables,

clasificando métodos como modelos interpretables, herramientas agnósticas y enfoques neuro-simbólicos. Este estudio se adhiere a la definición de Barredo et al. (2020) por exponer claramente la característica que se quiere agregar a la detección de acumulación de residuos, como es la transparencia y comprensibilidad en las explicaciones al modelo de red neuronal profunda Representative Points (REPPONTS), un modelo de “caja negra”.

### **Inteligencia Artificial Responsable**

La inteligencia artificial responsable se entiende como el conjunto de métodos que priorizan valores éticos, explicaciones comprensibles y la protección de la privacidad, seguridad y confianza de los usuarios (Goellner et al., 2024). La visión de XAI artificial explicable integra esos métodos mencionados anteriormente, buscando como objetivo agregar transparencia y comprensibilidad en los modelos de aprendizaje automático o aprendizaje profundo (Barredo et al., 2020). Por lo tanto, el prototipo de la investigación buscará dar explicaciones comprensibles y transparentes a los usuarios según la visión de XAI expuesta por Barredo (2020).

### **Conceptos Fundamentales**

Según Barredo et al. (2020), se señalan estos tres conceptos como claves sobre la XAI: interpretabilidad; es cuánto comprende el humano cómo las variables de entrada intervienen en las salidas; explicabilidad: incluye interpretabilidad, agregando la capacidad de justificar y comunicar decisiones entendibles; transparencia: nivel de exposición de la estructura interna del modelo, parámetros y datos que lo conforman. Igualmente, Páez (2025) expresa sobre estos tres conceptos lo siguiente: que la interpretabilidad se relaciona con cuánto el humano comprende el funcionamiento del modelo o cómo tomó la decisión, sea un experto o no, especialmente en los modelos complejos como las redes neuronales; en cuanto a la explicabilidad, busca reducir la opacidad algorítmica y fomentar la

confianza; y en la transparencia, la divide en epistémica al indicar que el modelo debe mostrar qué datos usa para tomar las decisiones, y la jurídica implica que cualquier persona tiene el derecho legal de acceder a una explicación sobre cómo la inteligencia artificial tomó la decisión.

Este estudio selecciona los conceptos fundamentales definidos por Barredo et al. (2020) aplicándolos al prototipo de visión computacional explicable-semántico, buscando no solo detectar si es o no una acumulación de residuos en la imagen, sino que también explicará qué factores se tomaron en cuenta al tomar la decisión; esto se realizará en la investigación mediante la funcionalidad de chat copiloto y la Application Programming Interface (API) de Generative Pre-trained Transformer 4 omni (GPT-4o), generando respuestas interpretables en lenguaje natural en formato texto y gráficos. Además, se buscará transparencia al mostrar el score de confianza de clasificación y las ubicaciones geoespaciales de la detección.

### **Taxonomía de Inteligencia Artificial Explicable**

La investigación de Barredo et al. (2020) propone dos taxonomías principales que son: taxonomía general para aprendizaje automático y para el aprendizaje profundo.

#### **Taxonomía General para Aprendizaje Automático**

Según Barredo et al. (2020) se organizan varios criterios que se presentan en el Anexo A, Tabla A2.

#### **Taxonomía Específica para Aprendizaje Profundo**

Según Barredo et al. (2020), se seleccionan para redes profundas las siguientes técnicas: atribución de características, donde se mide la contribución de cada parámetro

a la predicción; visualización de activaciones internas mediante gradientes o mapas de calor, que señalan las zonas que contribuyeron a la decisión; ejemplos y contrafactuales, usando casos similares para justificar la salida generada; y surrogados locales y globales, que es un método que usa modelos simplificados para interpretar el comportamiento general o una predicción concreta de la red neuronal profunda.

La investigación realizada por Schwalbe y Finzel (2024) generaliza los conceptos emitidos por Barredo et al. (2020) con métodos de inteligencia artificial explicable por tipo de explicación, ya sea visual o textual; según el nivel del usuario objetivo, si es experto técnico o usuarios finales; el momento en que se aplica la explicabilidad, si es antes, durante o después del entrenamiento; o el propósito de la explicación, para generar confianza, depuración del modelo, cumplimiento legal o normativo.

El prototipo implementado en esta investigación se adhiere a los conceptos de Barredo et al. (2020) aplicando los tres criterios de la siguiente forma: las explicaciones se generarán post-hoc en lenguaje natural, basadas en los resultados de las predicciones del modelo REPOINTS; en cuanto al tipo de explicaciones, el prototipo integra respuestas locales a preguntas como: “¿Qué corregimiento tiene más detecciones esta semana?” o explicaciones globales ante solicitudes como: “Resumen de detecciones”; en cuanto a la audiencia, las respuestas resultantes del chat inteligente están organizadas primero con las explicaciones para usuarios no técnicos, con narrativa explicativa gráfica, visual o textual, y segundo para técnicos mediante sentencias Structured Query Language (SQL) y tablas de datos.

En cuanto a la taxonomía de redes profundas, el modelo de inteligencia artificial que se usa en este modelo es REPOINTS con 20 millones de parámetros, debido a su carácter no interpretable de manera intrínseca se le considera de “caja negra” porque no es evidente cómo cada parte concreta lleva a una decisión final.

## Oportunidades de la Inteligencia Artificial Explicable

Las oportunidades que brinda la inteligencia artificial explicable, según Barredo et al. (2020), son: la construcción de confianza, que busca que los usuarios acepten la inteligencia artificial al entender sus decisiones; en cuanto al cumplimiento regulatorio, los modelos deben apoyar normativas que exijan transparencia; implementar mejoras a los modelos, permitiendo identificar sesgos o fallos internos. Además, favorecer la interacción humano-máquina para hacer los sistemas más comprensibles y nuevas métricas que impulsen el desarrollo de estándares para medir la calidad de la explicación.

Asimismo, según Schwalbe y Finzel (2024), las oportunidades de XAI incluyen la mejora de la confianza del usuario al ofrecer explicaciones claras para generar mayor aceptación y confianza; facilitación de la depuración de modelos con explicaciones que permitan identificar errores; cumplimiento normativo y legal que exigen transparencia en la toma de decisiones automatizadas; y en lo educativo, pueden ser utilizadas como herramientas pedagógicas para enseñar cómo funcionan los modelos.

La implementación del prototipo en esta investigación integra dichas oportunidades mediante las siguientes funciones: explicaciones claras de los factores tomados en cuenta en la detección por el modelo GPT-4o en lenguaje natural; además, buscando aportar en el cumplimiento de la Ley N.º 276 (2021) de Panamá con la detección automática de las acumulaciones de residuos en las calles en San Miguelito, Ciudad de Panamá; mejorando la interacción humano-máquina con la funcionalidad de chat en el prototipo con respuestas de forma textual, visual y gráficos; implementando métricas para medir la calidad de las explicaciones, con test automáticos incluidos en la herramienta.

## **Principales Desafíos**

En la investigación de Barredo et al. (2020) se indica que XAI enfrenta varios desafíos, como el equilibrio rendimiento–explicabilidad, que tiene como limitante que los modelos más simples suelen ser más comprensibles, pero menos precisos. También señala que la fidelidad de las explicaciones post-hoc conlleva el riesgo de que la explicación no refleje realmente cómo opera el modelo, aunque señala que la evaluación de explicaciones, dentro de la comunidad falta consenso sobre qué significa que una explicación sea “buena”. Igualmente, informa de otros desafíos, tales como: la complejidad de las redes profundas, con millones de parámetros, que dificulta la transparencia; la necesidad de que las explicaciones revelen discriminaciones ocultas; evitar dar demasiada información con datos sensibles; y tener en cuenta la diversidad de la audiencia para adaptarse al perfil del usuario ya sea técnico, regulador o ciudadano.

Los desafíos de XAI según Schwalbe y Finzel (2024) pueden ser: el equilibrio entre precisión y explicabilidad, ya que los modelos más precisos suelen ser menos interpretables, lo que genera un conflicto técnico; la adaptación al usuario objetivo, ya que las explicaciones deben ajustarse al nivel técnico del usuario; el riesgo de explicaciones engañosas o simplificadas, dado que técnicas pueden generar explicaciones que parecen claras pero no reflejan el funcionamiento interno del modelo; y limitaciones metodológicas, pues no todos los modelos permiten explicaciones útiles, e interpretables.

En este estudio se abordan dos retos: el primero de ellos es la fidelidad de las explicaciones post-hoc, añadiendo métricas de fidelidad de la información basadas en los datos para demostrar el apego a lo detectado; y el segundo es tener en cuenta la diversidad de la audiencia al organizar las respuestas en el chat del prototipo, con una sección para explicaciones en lenguaje natural, gráfico o visual para usuarios no técnicos y tablas con datos, además de la sentencia SQL utilizada para usuarios más expertos.

## **Modelos de Lenguaje de Gran Escala**

### **Definición**

Los modelos Large Language Model (LLM) son sistemas entrenados con un conjunto grande y estructurado de textos que generan texto natural simulando razonamiento (Astobiza, 2025). Según Kotz et al. (2024) llama al LLM como modelos que generan texto reflejando un mundo modelado basado en patrones léxicos. En esta investigación se utiliza la definición de Astobiza (2025), implementando en el prototipo la funcionalidad de Chat con GPT-4o como motor de consulta en lenguaje natural, permitiendo solicitudes en español y simulando razonamiento con respuestas explicativas de las predicciones realizadas, aportando más que solo detección de residuos.

### **Representación Semántica**

Según Astobiza (2025), expresa que, si la pregunta o el concepto es concreto y contextualizado, el LLM puede responder bien porque sabe correlacionar palabras a datos estructurados; en cambio, si el concepto es abstracto o muy teórico, el modelo puede generar una respuesta, pero no representaría el entendimiento a profundidad de ese concepto. Del mismo modo, Kotz et al. (2024) expresa que los LLM no comprenden el significado profundo, sino que simulan coherencia lingüística. En el prototipo, se supera en parte esa limitación, basando las respuestas del modelo en los datos resultantes de la consulta a la base de datos y manteniendo los parámetros temperatura, top, semilla con valores fijos durante los prompts o reprompts. Toda esta información mencionada funciona como ancla a las respuestas, minimizando que el modelo de lenguaje de gran escala represente el entendimiento a profundidad.

## **Construcción de Teorías**

Teorizar implica formular hipótesis, conceptualizar relaciones y abstraer conocimiento; los modelos de lenguajes de gran escala pueden aportar en este proceso, sin sustituirlo (Astobiza, 2025). Aunque Kotz et al. (2024) no aborda directamente el tema de la construcción de teorías, más bien sugiere que los modelos representan estructuras del conocimiento sin reemplazar totalmente la construcción. Por consiguiente, el prototipo implementado en esta investigación toma la teoría expuesta por Astobiza (2025) y la aplica en el chat con LLM GPT-4o, formulando hipótesis sobre patrones en la acumulación de residuos en el distrito de San Miguelito, Ciudad de Panamá, sin reemplazar por completo el proceso técnico ya establecido.

## **Limitaciones**

Los modelos de lenguajes de gran escala no validan hipótesis científicas de manera independiente ni manejan bien conceptos altamente profundos (Astobiza, 2025); el autor Kotz et al. (2024) confirma estas limitaciones, indicando que los LLM tienen restricciones cognitivas y propone como alternativa los Large World Models (LWMs). En esta investigación se señala que el prototipo es una herramienta de detección y explicabilidad post-hoc que apoya la interpretación, sin embargo, la validación de los resultados se realiza mediante análisis humano o métricas técnicas, apoyándose en la teoría expuesta de Astobiza (2025).

## **Text-to-SQL Basado en Modelos de Lenguaje de Gran Escala**

### **Definición**

Según Hong et al. (2024), LLM-based Text-to-SQL es la tarea de convertir preguntas en lenguaje natural en consultas SQL ejecutables sobre una base de

datos usando modelos LLM; a la par, Mohammadjafari et al. (2025) lo conceptúa como traducir consultas en lenguaje natural de usuarios no expertos a crear consultas de bases de datos relacionales. La implementación del prototipo en este estudio transforma semánticamente la consulta en lenguaje natural a consultas SQL mediante el modelo GPT-4o, complementando las definiciones de Hong et al. (2024) y Mohammadjafari et al. (2025).

## **Objetivo**

El objetivo de transformar consultas en lenguaje natural, según Hong et al. (2024), es permitir que usuarios sin conocimientos técnicos profundos de consultas que interactúen con bases de datos, haciendo natural la interfaz con la base de datos. De igual manera, Mohammadjafari et al. (2025) comentan que busca facilitar la consulta de datos sin necesidad del conocimiento profundo de la sintaxis técnica de las consultas. Este estudio se alinea con estos objetivos complementarios, permitiendo con el prototipo consultas como “resumen de las detecciones de hoy” y a continuación el LLM interpreta la sentencia correspondiente sin romper la estructura semántica.

## **Componentes Clave**

Según Hong et al. (2024), los componentes principales en sistemas Text-to-SQL con los modelos LLM son: comprensión de la pregunta en lenguaje natural; entendimiento del esquema de la base de datos, de tablas, columnas y relaciones, y la generación de la consulta propiamente dicha. Igualmente, Mohammadjafari et al. (2025) define componentes similares como la comprensión del lenguaje natural, enlace con esquemas y la generación de las instrucciones de consulta. Cabe resaltar que agrega dos componentes más, como el manejo de la complejidad con consultas cross-domain y el uso de Retrieval-Augmented Generation (RAG) para mejorar el contexto. Los tres componentes abordados por Hong et al. (2024) encajan perfectamente en este prototipo, comprendiendo consultas en lenguaje

natural por medio del chat con GPT-4o; entendiendo el esquema de la base de datos, ya que se envía el esquema de la base de datos y la información en cada prompt; y, por medio de parsing semántico, transforma el lenguaje natural a sintaxis de consultas estructuradas semánticamente para base de datos.

## Métricas de Evaluación

Los autores Hong et al. (2024) y Mohammadjafari et al. (2025) coinciden en señalar cuatro métricas principales de evaluación Text-to-SQL: Component Matching (CM) y Exact Matching (EM), que están orientadas a medir la correspondencia estructural entre la consulta generada por el modelo y la referencia humana; y Execution Accuracy (EX) junto a Valid Efficiency Score (VES), que se enfocan en la validez semántica y la eficiencia en la ejecución de la base de datos. Por lo tanto, dado que uno de los objetivos específicos de este estudio es valorar la fidelidad semántica de las respuestas del modelo, se selecciona la métrica EX, expuesta por Hong et al. (2024), para evaluar la fidelidad semántica del modelo LLM a nivel de consultas a base de datos, asimismo utilizar el criterio binario de que cada ítem cuenta como “aprobado” sólo si la sentencia generada coincide con el esperado en las mismas columnas y mismo conjunto de filas.

## Fases en Proyectos de Inteligencia Artificial

Las fases en proyectos de XAI comentadas por Molnar (2019) son las siguientes:

1. Diseño y entrenamiento: aplicando la inteligencia artificial explicable desde el diseño o con técnicas post-hoc.
2. Evaluación de generalización: deben centrarse en que las explicaciones se mantienen coherentes y comprensibles.
3. Evaluación en campo: probar la utilidad del modelo en contextos reales para aumentar la confianza.

Adicionalmente, el autor Barredo et al. (2020) describe fases aplicadas a la inteligencia artificial explicable a continuación:

1. Selección del método: según la naturaleza del problema y del objetivo a alcanzar, se define si se requiere interpretabilidad total o explicaciones locales.
2. Aplicación del método explicativo: se aplican métodos como Local Interpretable Model-agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP) o Gradient-weighted Class Activation Mapping (GRAD-CAM), u otros métodos.
3. Evaluación de fidelidad: medir si las explicaciones del modelo son comprensibles, fieles y útiles para tomar decisiones.

Este estudio selecciona las fases expuestas por el autor Barredo et al. (2020) por tener mayor acercamiento al objetivo de la investigación de implementación de un prototipo con visión computacional explicable-semántico, estas etapas descritas se verán aplicadas en la sección de procedimientos, resultados, análisis de datos y conclusiones.

## **Bases Legales**

### **Definición de Residuo**

De acuerdo Ley N.<sup>o</sup> 276 (2021) de Panamá, un residuo se define como parte o porción del resultado de actividades humanas que pueden ser utilizadas o no de aprovechamiento para uno nuevo. En el contenido de esta investigación utilizará la palabra residuos para referirse a desperdicio orgánico o no orgánico o sólido que comúnmente se conoce como basura, exceptuando los casos cuando se usa para citar a otros autores que usan palabras como basurales y similares.

### **Responsabilidades de la Autoridad de Aseo Urbano y Domiciliario**

Según el artículo 5 de la Ley N.<sup>o</sup> 276 (2021) de Panamá, el Estado, a través de la Autoridad de Aseo Urbano y Domiciliario, es responsable de garantizar los

servicios de limpieza urbana y gestión integral de residuos domiciliarios, peligrosos, comerciales y especiales.

Este estudio adoptará como base la Ley N° 276 (2021) de Panamá en la definición de residuo y la responsabilidad de la Autoridad de Aseo Urbano y Domiciliario en la gestión de residuos, proponiendo como valor añadido la detección automática y explicativa de acumulación de residuos en las calles con visión computacional. En el Anexo A, Tabla A3 aparece el resumen de las teorías seleccionadas en esta investigación con el respectivo aporte.

### **Términos básicos**

Aprendizaje automático: "es una rama de la inteligencia artificial que busca un programa de computador aprenda de un conjunto de datos con los cuales se entrena, y buscará identificar un patrón" (Pineda, 2022, pág. 36).

Archivos Keyhole Markup Language (KML): "es un formato de archivo que se utiliza para mostrar datos geográficos en un navegador de Google Earth, como Google Earth. KML utiliza una estructura basada en etiquetas con elementos y atributos anidados" (Google Developers, s. f.-a, párr. 1).

Geographical JavaScript Object Notation (GEOJSON): "es un formato de intercambio de datos geoespaciales de código abierto que representa entidades geográficas sencillas y sus atributos no espaciales" (Esri, s. f.-a, párr. 1).

Geographic Information System (GIS): "son sistemas informáticos que producen visualizaciones conectadas de datos geoespaciales" (Jonker, 2023, párr. 1).

Google Earth Pro: "es un sistema de información geográfica que muestra un globo terráqueo virtual que permite visualizar múltiple cartografía, basado en imágenes satelitales" (IES Cerro del Viento, 2023, párr. 2).

Generative Pre-trained Transformer (GPT): “es un modelo de inteligencia artificial dedicado a tareas de procesamiento automatizado de lenguaje natural” (DataScientest, 2024a, párr. 1).

Gradient-weighted Class Activation Mapping (GRAD-CAM): “El método Grad-CAM es una técnica popular de visualización que es útil para comprender cómo se ha conducido a una red neuronal convolucional a tomar una decisión de clasificación” (DataScientest, 2021, párr. 9).

Inferencia en inteligencia artificial: “La inferencia consiste en poner en práctica lo que la IA ha aprendido en el entrenamiento” (Neuroons, 2021, párr. 6).

Inteligencia Artificial: “la inteligencia artificial es un futurista campo de estudio y desarrollo que se centran la creación de sistemas y programas capaces de realizar tareas que requieren la participación de la inteligencia humana” (Zeron, 2023, pág. 1).

LandingLens: “permite entrenar sistemas de visión basados en inteligencia artificial para identificar objetos, patrones o defectos sin necesidad de conocimientos avanzados” (C de Comunicación, 2023, párr. 2).

Large World Models (LWM): “buscan dotar a la IA de una inteligencia espacial similar a la humana, con modelos que pueden percibir, generar e interactuar con entornos tridimensionales” (Expansión, 2025, párr. 6).

Local Interpretable Model-agnostic Explanations (LIME): “es un algoritmo que puede explicar las predicciones de cualquier clasificador o regresor de una manera fiel, aproximándolo localmente con un modelo interpretable” (Gonzalez, 2023, párr. 2).

Matriz de confusión para aprendizaje automático: “es un instrumento tecnológico que sirve para calcular el rendimiento sobre un modelo de clasificación definido” (DataScientest, 2024b, párr. 3).

Memex: “permite a los desarrolladores construir software de manera más rápida e intuitiva. Con su interfaz basada en lenguaje natural y sus capacidades de inteligencia artificial” (Toolify, 2025, párr. 1).

Modelo: “es el conjunto de parámetros y la estructura necesarios para que un sistema realice predicciones.” (Google, s.f.-c, párr. 79)

MMDetection: “un conjunto de herramientas especializado en detección de objetos” (Parra, 2023, pág. 38).

Overpass Turbo: “es una herramienta web de minería de datos para OpenStreetMap (OSM). Overpass turbo puede ejecutar cualquier tipo de consulta de datos y muestra los resultados en un mapa interactivo” (Morales, 2018, párr. 1).

Precisión para aprendizaje automático: “La precisión es la proporción de todas las clasificaciones positivas del modelo que realmente son positivas” (Google Developers, s.f.-b, párr. 19).

Prototipo: “Un primer modelo de un producto o servicio que se lleva intención de testear; o bien, lanzar al mercado” (Universidad Europea, 2024, párr. 1).

Python: “Python es un lenguaje dinámico, interpretado (compilado en bytecode). No hay declaraciones de tipo de variables, parámetros, funciones o métodos en el código fuente” (Google Developers, s.f., párr. 3).

Quantum Geographic Information System (QGIS): “es un cliente SIG de escritorio amigable de código abierto donde se puede visualizar, administrar, editar, analizar datos y componer mapas” (OSGeo, 2015, párr. 1).

Tasa de verdaderos positivos (Recall): “la proporción de todos los positivos reales que se clasificaron correctamente como positivos” (Google Developers, s.f., párr. 9).

Redes Neuronales Artificiales: "es una solución de software que utiliza algoritmos de aprendizaje automático para «imitar» las operaciones del cerebro humano. " (Domínguez, 2025, pág. 1).

Representative Points (REPPONTS): "es un detector de dos etapas que utiliza un conjunto adaptativo de puntos para la detección de diferentes objetos orientados" (Torregrosa, 2023, pág. 19).

SHapley Additive exPlanations (SHAP): "Son una forma de explicar la salida de cualquier modelo de machine learning. Utiliza un enfoque de teoría de juegos que mide la contribución de cada jugador al resultado final" (Wan, 2024, párr. 6).

Structured Query Language (SQL): "Es un lenguaje de programación estandarizado que se utiliza para administrar bases de datos relacionales y realizar diversas operaciones con los datos que contienen" (Sirkin, 2021, párr. 1).

SQLite: "es una herramienta de software libre, que permite almacenar información en dispositivos empotrados de una forma sencilla, eficaz, potente, rápida y en equipos con pocas capacidades de hardware" (Rommel, s. f., párr. 3).

Redes neuronales transformer: "El modelo de transformador es un tipo de arquitectura de redes neuronales que destaca en el procesamiento de datos secuenciales, asociado sobre todo a los modelos de lenguaje de gran tamaño (LLM)" (Striker y Bergmann, 2025, párr. 1).

Temperatura: "Controla la aleatoriedad: los valores más bajos producen resultados más deterministas, mientras que los valores más altos introducen más creatividad" (Tuychiev, 2025, párr. 12).

Teselas en KML: "Equivalen a un esquema de ordenamiento en teselas o de caché de imágenes" (Esri, s. f.-b, párr. 9).



## **METODOLOGÍA**

### **Métodos de Investigación**

Los métodos de investigación se dividen según Tarrillo et al. (2024) en: inductivo, deductivo, inductivo-deductivo, hipotético-deductivo, histórico, analítico, sintético. En este estudio se seleccionó el hipotético-deductivo por iniciar con la hipótesis de que la implementación del prototipo con visión computacional explicable-semántica detectará de forma fiable la acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá, y se deducirán las conclusiones confrontándolas con la realidad simulada en el experimento.

### **Tipos de Investigación**

Los tipos de investigación según Tarrillo et al. (2024) se dividen principalmente por la finalidad del estudio, el enfoque, el diseño y nivel. Este estudio ha seleccionado lo siguiente: finalidad del estudio: la finalidad seleccionada de este estudio será la aplicada, ya que aportará a las autoridades municipales de San Miguelito, Ciudad de Panamá, mayor facilidad en la toma de decisiones del proceso de recolección de basura mediante la implementación del prototipo con visión computacional para detectar acumulación de residuos en las calles.

Además, en el enfoque: la investigación estará enfocada en cuantitativa, debido a que todas las mediciones y validaciones del modelo de visión computacional explicable se realizarán con resultados numéricos y porcentajes; diseño: este estudio ha escogido como diseño el cuasiexperimental bajo un entorno simulado, donde se realizaría la intervención tecnológica con los métodos de inteligencia artificial explicable y se compararía la precisión general del modelo de visión computacional en la detección de acumulación de residuos; nivel: esta investigación tendrá como nivel el aplicativo, por realizar la implementación del

prototipo con visión computacional explicativa para detectar acumulaciones de basura.

## **Variables**

Las variables, según Tarrillo et al. (2024), son atributos o propiedades que tienen la capacidad de cambiar de valor y que pueden ser medidas en un estudio, siendo componentes principales que ayudan a los investigadores a comprender y analizar los eventos en respuesta a diferentes condiciones. Además, según los autores Tarrillo et al. (2024), se clasifican, según su naturaleza, en variables categóricas con dos subtipos: nominales u ordinales, y variables cuantitativas divididas en discretas y continuas. Igualmente, las clasifican según su relación con el estudio en variables dependientes e independientes.

En este estudio se determinan dos variables y según su naturaleza: acumulación de residuos en las calles, de naturaleza cuantitativa continua, ya que se va a contar la cantidad de residuos detectados y cálculos porcentuales de distribución por corregimiento que pueden ser cualquier número en una escala definida; en cuanto a la variable de prototipo de visión computacional explicable-semántico, por su naturaleza, se clasifica como continua, porque el valor de la métrica de precisión o fidelidad puede ser cualquiera en un rango.

En cuanto a la relación con el estudio, la variable dependiente es la acumulación de residuos en las calles, que está definida y derivada del problema central de este estudio, que es la gran acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá. La variable independiente es el prototipo con visión computacional explicable-semántico, ya que se evaluará la precisión en detección y la fidelidad explicativa de la variable dependiente durante el estudio.

## **Operacionalización de Variables**

La operacionalización de las variables es la traducción de cada variable de la definición conceptual en el estudio en indicadores y dimensiones medibles (Tarrillo et al., 2024). Además, los autores Tarrillo et al. (2024) explican que las dimensiones de las variables son los componentes específicos en que se descompone la variable para observarla con mayor detalle, y que los indicadores se definen como señales concretas medibles para evaluar cada dimensión, a continuación, en el Anexo B, Tabla B1 el estudio presenta la operacionalización de variables.

## **Población y Muestra**

De acuerdo con Tarrillo et al. (2024), “la población es el conjunto de elementos o individuos que comparten una característica común y son objeto de estudio. Representa el universo completo sobre el cual se quieren hacer inferencias” (pág. 79). En este estudio, la población comprende todas las imágenes satelitales del distrito de San Miguelito, Ciudad de Panamá, que representan las posibles acumulaciones de residuos en espacios abiertos, como calles, aceras y lotes baldíos. Por su parte, “la muestra es un subconjunto representativo de la población que se selecciona para participar en la investigación. La idea es que los resultados obtenidos de la muestra puedan generalizarse a la población de forma ampliada” (Tarrillo et al., 2024, p. 79).

En esta investigación, la muestra se compone de dos grupos en períodos específicos: abril de 2023 para el entrenamiento del modelo de visión computacional subdivididas en 70% grupo train, 20% grupo dev y 10% para grupo test; y enero de 2024 con 2227 imágenes para la validación de la precisión general del prototipo, coincidiendo con la fecha de la alerta sanitaria emitida por el Ministerio de Salud de Panamá en enero de 2024, que se señaló en el planteamiento del problema.

En cuanto a las imágenes seleccionadas para entrenamiento solo tienen un criterio de exclusión, que es balancear la cantidad de imágenes positivas y negativas para equilibrar las inferencias del modelo, minimizando el sesgo, esta distribución de la muestra entre negativos y positivos es de 834 imágenes en LandingLens, en el Anexo B, Figura B1 muestra la distribución alcanzada. El conjunto de validación de enero 2024 que se usará para calcular la matriz de confusión su distribución por teselas es detallado en el Anexo B, Tabla B2.

### **Técnica e Instrumentos**

Según Tarrillo et al. (2024), la técnica representa métodos científicos que pueden ser de observación, encuesta y experimental, que utiliza el investigador para obtener datos relevantes del estudio realizado. Los instrumentos, según señalan los autores, son las herramientas concretas, sean físicas o virtuales, que se utilizarán en el proceso de recolección de datos. En este estudio, el enfoque seleccionado es el cuantitativo en ambas variables y las técnicas e instrumentos son los siguientes: para la variable dependiente acumulación de residuos en las calles, se empleó la técnica de observación estructurada, utilizando como instrumento una ficha de observación derivada del antecedente de Cañete (2024), en el Anexo B, Figura B2 se presenta la ficha. Esto permitió registrar de forma sistemática la cantidad de residuos en cada tesela de imagen de 1280 x 720 píxeles para tener visibilidad de la cantidad y la distribución espacial de la acumulación de residuos en el distrito de San Miguelito, Ciudad de Panamá.

De igual forma, los datos recolectados aportan las etiquetas reales que se compararán con las predicciones automáticas por el prototipo con visión computacional explicable-semántico. En cuanto a la variable independiente del prototipo con visión computacional explicable-semántico, como técnica se utilizó la evaluación post-hoc de explicaciones multimodales con pruebas automáticas dentro del prototipo, y como instrumento se presenta un informe de evaluación como se muestra en el Anexo B, Figura B3 dividido en dos bloques: el primero, la

precisión general del modelo de visión computacional calculada con la comparación entre etiquetas manuales de la ficha de observación y predicciones automáticas con toda la población de imágenes satelitales de enero de 2024; el segundo, la fidelidad de explicabilidad post-hoc que se mide a través de la métrica Execution Accuracy (EX) explicada por Hong et al. (2024) para evaluar la fidelidad semántica a nivel de consultas a base de datos, y la métrica virtual-textual Similarity (VTS) propuesta por Li et al. (2024) para valorar la fidelidad explicativa a nivel multimodal, además se utilizará el porcentaje más alto de las referencias de ejecuciones anteriores de la misma métrica VTS según la recomendación de los autores.

## **Procedimientos**

De acuerdo con Tarrillo et al. (2024), los procedimientos de una investigación representan los pasos organizados que permiten aplicar las técnicas y los instrumentos previamente definidos, con el objeto de garantizar la validez y confianza en la recolección de los datos. En este estudio, la estructura de los procedimientos está en correspondencia con los objetivos específicos que son contabilizar residuos, construir el prototipo, medir el rendimiento general y evaluar la fidelidad explicativa-semántica del prototipo. Cada objetivo se complementa con las fases de inteligencia artificial explicable propuestas por Barredo et al. (2020), que son la selección del método explicativo, la aplicación del modelo y la evaluación de fidelidad.

### **Construir el Prototipo: Fase de Diseño y Entrenamiento**

La fase de diseño y entrenamiento se divide en los siguientes procedimientos específicos: la selección de la explicabilidad, la obtención de límites administrativos, el historial de imágenes, la captura de imágenes satelitales, el etiquetado y la preparación del conjunto de imágenes, el entrenamiento del modelo y el diseño de la aplicación web.

## **Selección de Explicabilidad**

Según Barredo et al. (2020), en la fase del diseño se debe seleccionar en qué lugar se van a dar las explicaciones; estas pueden ser intrínsecas o post-hoc. Esta investigación seleccionó la explicabilidad post-hoc, que sería aplicada después de entrenar el modelo Representative Points (REPPONTS) en la plataforma LandingLens.

## **Obtención de Límites Administrativos**

Inicio de la búsqueda de los límites administrativos de cada corregimiento que componen el distrito de San Miguelito, Ciudad de Panamá, mediante la plataforma Overpass Turbo, realizando consultas Overpass como el ejemplo del Anexo B, Figura B4, para el corregimiento Omar Torrijos. Cada resultado en Overpass se exportó a archivos de Keyhole Markup Language (KML), que representan gráficamente áreas completas de interés en el mapa de Google Earth Pro. Además, estos archivos se deben segmentar las áreas completas en teselas de 150 metros y zoom 213 para ser usados como guías al realizar capturas satelitales mediante la funcionalidad de creador de películas en Google Earth Pro, en el Anexo B, Figura B5 se muestra el contenido del archivo.

## **Historial de Imágenes**

Según lo indicado en la sección de población, las imágenes de entrenamiento serían de la penúltima fecha disponible con imágenes satelitales. Utilizando la opción de historial de imágenes de Google Earth Pro, sería abril de 2023 y en validación con fecha enero 2024.

## **Captura de Imágenes Satelitales**

Este paso se llevó a cabo cargando cada archivo KML segmentado por el paso anterior, además de realizar estas acciones: seleccionar la tesela 1 con doble clic, ajustar la vista norte desde arriba a cero, hacer zoom a 213, quitar la selección de la tesela y ejecutar la funcionalidad creadora de películas seleccionando el viaje que se creó con la segmentación. Es preciso señalar que el proceso anterior se ejecutó en dos ocasiones: la primera con abril 2023 y la segunda con fecha enero 2024 en la herramienta Google Earth Pro. En el Anexo B, Figura B6 se presenta los parámetros que se usaron al momento de generar la captura masiva de imágenes satelitales por teselas en cada fecha.

## **Etiquetado y Preparación del Conjunto de Imágenes**

Este procedimiento se efectuó en la plataforma LandingLens seleccionando una muestra de las imágenes de abril de 2023, con 632 imágenes en total equilibradas entre positivas y negativas, según lo acordado en la sección de población y muestra. Seguidamente, se inicia el proceso de etiquetado por cada imagen cargada en LandingLens, señalando la ubicación de la acumulación de residuos y asignando automáticamente la clase 'residuos'. Si no hay, se asignará la clase 'no class', se puede ver en el Anexo B, Figura B7, un ejemplo de etiquetado en imágenes con residuos.

## **Entrenamiento del Modelo**

Este paso fue realizado en tres ocasiones, basándose inicialmente en 632 imágenes. En el Anexo B, la Tabla B3 muestra los hiperparámetros utilizados, la augmentación asignada y la distribución de las imágenes en el primer entrenamiento. En el segundo entrenamiento se incluyen 177 imágenes positivas del conjunto de imágenes público AerialWaste, utilizando los mismos

hiperparámetros del entrenamiento anterior, y la distribución del conjunto de imágenes queda como se muestra en el Anexo B, Tabla B4. En el tercer y último entrenamiento, se eliminaron las imágenes del conjunto de datos público AerialWaste porque provocaron una caída en la precisión y el recall del modelo. En consecuencia, se implementa como mejora la generación de imágenes aumentadas en un Jupyter Notebook local con los parámetros en Python presentados en el Anexo B, Figura B8. Las imágenes con augmentación se agregaron al grupo de entrenamiento de forma equilibrada, y la distribución final es como se muestra en el Anexo B, Tabla B5.

### **Diseño de la Aplicación Web**

La aplicación web se diseñó mediante el asistente de codificación Memex, usando el prompt detallado en el Anexo B, Figura B9. Igualmente, como requisitos previos, se suministró a Memex la Application Programming Interface Key (API KEY) de OpenAI, el modelo Representative Points (REPPONTS) descargado desde LandingLens optimizado con Open Visual Inference and Neural Network Optimization (OPENVINO). Además, por medio de prompts se le agregaron funciones para descargar imágenes desde Google Drive y simular orígenes remotos; sumado a ello, se agregaron funcionalidades de exportación de las detecciones a archivos KML y archivos Comma Separated Values (CSV). El resultado final de todo el prototipo está en el Anexo B, Figura B10 hasta el Anexo B, Figura B14.

### **Contabilizar la Acumulación de Residuos: Fase de Diseño y Entrenamiento**

En esta fase, el procedimiento realizado en la investigación fue completar los ítems de la ficha de observación del Anexo B Figura B2, que son: nombre de imagen, corregimiento, cantidad de detecciones por cada imagen satelital generada por tesela de forma automática en la acción específica de captura de imágenes satelitales con fecha de enero de 2024 de la fase anterior. Al terminar,

el Excel con los datos recolectados se transformará en un gráfico de anillo con la distribución porcentual por corregimiento y en otro gráfico de tipo columna por cantidad de detección por corregimiento. Esto cumplirá con las dimensiones de la variable dependiente al dar visibilidad de la cantidad y distribución espacial de los residuos por corregimiento en el distrito de San Miguelito, Ciudad de Panamá. En el Anexo B, Figura B15 muestra un ejemplo de los gráficos.

### **Medir el Rendimiento General del Modelo: Fase de Evaluación de Generalización**

En esta fase del estudio, el procedimiento ejecutado es la elaboración de un informe de evaluación como se presenta en el Anexo B, Figura B3 con el primer bloque, manteniendo como alcance la dimensión de rendimiento del modelo de la variable independiente. Dentro del informe se presenta el indicador de precisión general a partir de la matriz de confusión con los datos de etiquetado real y las predicciones realizadas por el prototipo de visión computacional, las acciones específicas para crear el indicador están detalladas a continuación son: ejecución dentro del prototipo del proceso de predicción masiva de las imágenes en Google Drive simulando orígenes remotos.

Luego, exportar las predicciones a archivos Comma Separated Values (CSV) desde el prototipo; seguidamente, subir los archivos CSV de las predicciones del prototipo y el archivo CSV de las etiquetas reales al Jupyter Notebook local. Por consiguiente, se calcula la matriz de confusión y la precisión general del modelo con Python dentro del Jupyter Notebook Local. En el Anexo B, la Figura B16 muestra la matriz de confusión y métricas.

### **Evaluar la Fidelidad Explicativa-Semántica: Fase de Evaluación Simulada.**

En esta etapa del estudio, se terminó de documentar el informe de evaluación que se presenta en el Anexo B, Figura B3 en su segundo bloque, conservando como alcance la dimensión de fidelidad explicativa de la variable

independiente del modelo, aplicando las métricas EX de la taxonomía de Hong et al. (2024) para medir la explicabilidad semántica a nivel de consultas a la base de datos, esta métrica usa como insumo un archivo en formato CSV con 45 filas de ejemplos de consultas de referencia y preguntas en lenguaje natural creados en esta investigación que la métrica usará durante su proceso; en cuanto a la métrica VTS de Li et al. (2024) mide la fidelidad semántica a nivel de inteligencia artificial explicable multimodal.

Además, para ejecutar esta medida se crearon registros en base de datos con equivocaciones intencionales muy drásticas cada 5 ejemplos de 40 ejemplos en total de explicaciones creadas aleatoriamente. Ambas métricas serán ejecutadas mediante el chat copiloto del prototipo, que tiene como funcionalidad ejecutar pruebas automatizadas de las métricas descritas anteriormente escribiendo “metricasllms” para EX y “metricasllms vts mbr” para VTS, incluyendo un modal para ir a una página del prototipo con los resultados, en el Anexo B, Figura B17 se presentan los prompts usados en Memex para construir ambas métricas.

## **ANÁLISIS DE DATOS Y RESULTADOS**

El análisis de datos se entiende como el proceso de interpretar y extraer conclusiones a partir de toda la información recolectada en la investigación, aplicando técnicas estadísticas, matemáticas y computacionales; buscando significado a los datos (Tarrillo et al., 2024). De igual forma, Tarrillo et al. (2024) definen la sección de resultados como el espacio donde el investigador presenta de manera objetiva los hallazgos obtenidos por la aplicación de los instrumentos. En coherencia con ello, el análisis de datos y los resultados en este estudio serán organizados en las fases ya establecidas en los procedimientos y vinculadas a los objetivos específicos. Esto asegura continuidad metodológica en el diseño, ejecución, interpretación y exposición objetiva de los resultados.

### **Análisis de datos**

#### **Construir el Prototipo: Fase de Diseño y Entrenamiento**

En esta fase se analizaron los registros derivados de las predicciones realizadas por el modelo durante los tres entrenamientos ejecutados en LandingLens, en cada ocasión se generaron indicadores de rendimiento estándar en proyectos de inteligencia artificial con visión computacional como los siguientes: precisión, que mide qué tan buenas son las predicciones positivas del modelo, la fórmula completa aparece en el Anexo B, Figura B18; sensibilidad (recall), mide qué proporción de los positivos son detectados, en el Anexo B, Figura B19 se muestra la fórmula usada; Mean Average Precision (MAP), es la exactitud media de precisión del modelo de qué tan bien detecta y localiza un objeto en las imágenes, en el Anexo B, Figura B20 se encuentra la fórmula utilizada; finalmente, la matriz de confusión, que funciona como tabla de estadística descriptiva que cruza las etiquetas reales y las predichas, en el Anexo B, Figura B21 se muestra la fórmula utilizada. Estas métricas aportan una valoración objetiva del rendimiento del prototipo en su etapa inicial de entrenamiento.

## **Contabilizar la Acumulación de Residuos: Fase de Diseño y Entrenamiento**

En esta etapa, el análisis se enfocó en la cuantificación sistemática de las acumulaciones de residuos en las imágenes satelitales de enero de 2024 generadas desde Google Earth Pro, dicho conteo genera la línea base del fenómeno de estudio permitiendo representar la distribución espacial de los residuos en el distrito de San Miguelito en gráficos comprensibles. Este análisis ofreció una referencia comparativa para la evaluación del modelo.

## **Medir el Rendimiento General del Modelo: Fase de Evaluación de Generalización**

El análisis de los datos en esta fase se centró en la elaboración de la matriz de confusión, comparando las etiquetas reales recolectadas con la ficha de observación y las predicciones generadas por el prototipo sobre la muestra de imágenes de enero 2024. De esta tabla se derivaron indicadores como la precisión general del prototipo que mide el porcentaje de aciertos totales del modelo, siguiendo el aporte señalado de la investigación que elaboraron Cárdenas et al. (2024) de utilizar dicha métrica con valor principal de medición, igualmente se incluye la fórmula utilizada en el Anexo B, Figura B22. Por lo tanto, la obtención de esta métrica ofrece una evaluación robusta de la capacidad del modelo para generalizar más allá del conjunto de entrenamiento.

## **Evaluar la Fidelidad Explicativa-Semántica: Fase de Evaluación Simulada**

El análisis de esta fase se dirigió a la fidelidad de las explicaciones generadas por el módulo post-hoc de explicabilidad integrado en el prototipo. Para ello, se implementaron dos métricas que se complementan: Execution Accuracy (EX) descrita por Hong et al. (2024), que mide la fidelidad semántica de las consultas a nivel de base de datos generadas por el Large Language Model (LLM) en comparación con los resultados de referencia en el insumo creado en la sección

de procedimientos, en el Anexo B, Figura B23 se muestra la fórmula utilizada en la métrica; y Visual-Textual Similarity (VTS) planteada por Li et al. (2024), que evalúa la coherencia entre las explicaciones textuales del LLM y las anotaciones visuales resaltadas en las imágenes, en el Anexo B, Figura B24 se presenta la fórmula utilizada en la medición. Estas métricas permiten analizar la utilidad y consistencia de las explicaciones ofrecidas, asegurando que el prototipo genere respuestas comprensibles y confiables al usuario.

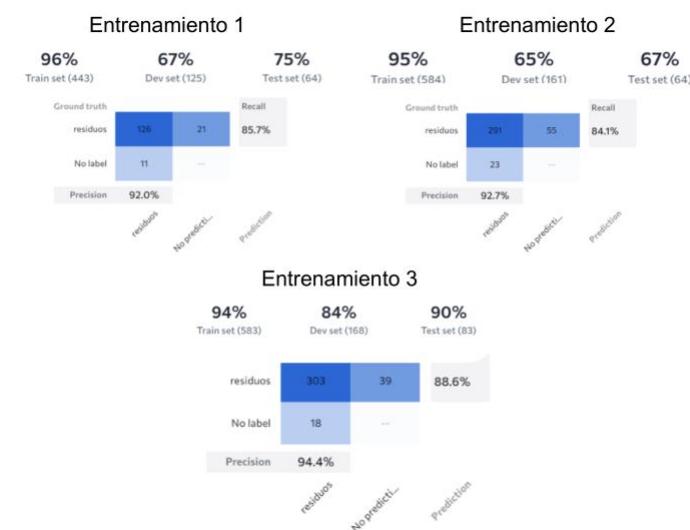
## Resultados

### Construir el Prototipo: Fase de Diseño y Entrenamiento

En la Figura B25 se presentan los resultados de los tres entrenamientos etiquetados en la imagen como “Entrenamiento 1”, “Entrenamiento 2” y “Entrenamiento 3” del modelo de visión computacional realizados en la plataforma LandingLens, que incluyen los valores de precisión general en la primera parte y en la segunda la matriz de confusión.

**Figura B25**

*Matriz de Confusión y Métricas Estándar*

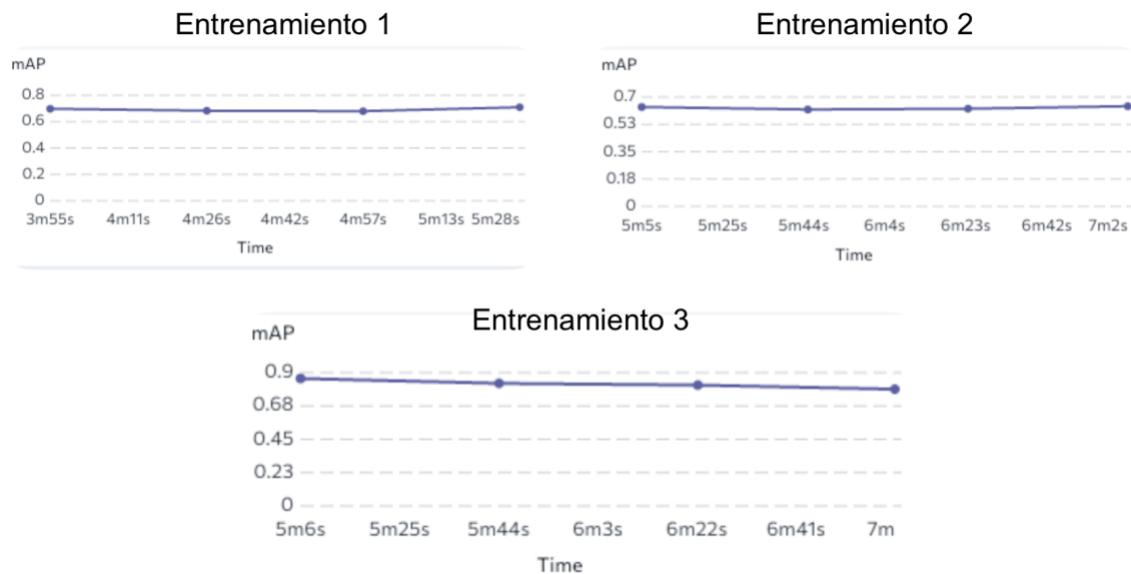


*Nota.* Captura de pantalla de indicadores de rendimiento en LandingLens (2025).

A continuación, se muestran en la Figura B26 los resultados de la métrica Mean Average Precision (MAP) en la plataforma LandingLens, donde el eje x es el tiempo que le toma al modelo en llegar a su punto más alto o bajo durante el entrenamiento y el eje y son los valores de MAP durante el entrenamiento.

**Figura B26**

*Resultados Mean Average Precision (MAP) Durante los Entrenamientos*



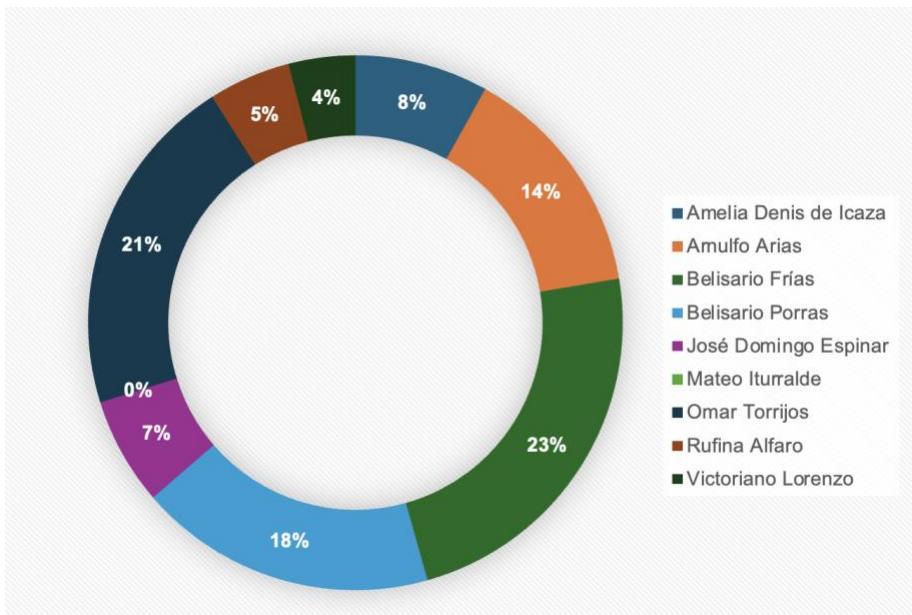
*Nota.* Captura de pantalla del gráfico en LandingLens (2025).

#### **Contabilizar la Acumulación de Residuos: Fase de Diseño y Entrenamiento**

En la Figura B27 el gráfico circular indica las distribuciones porcentuales por corregimiento obtenidas del conteo sistemático con la ficha de observación.

**Figura B27**

*Distribución Porcentual de Acumulaciones de Residuos*

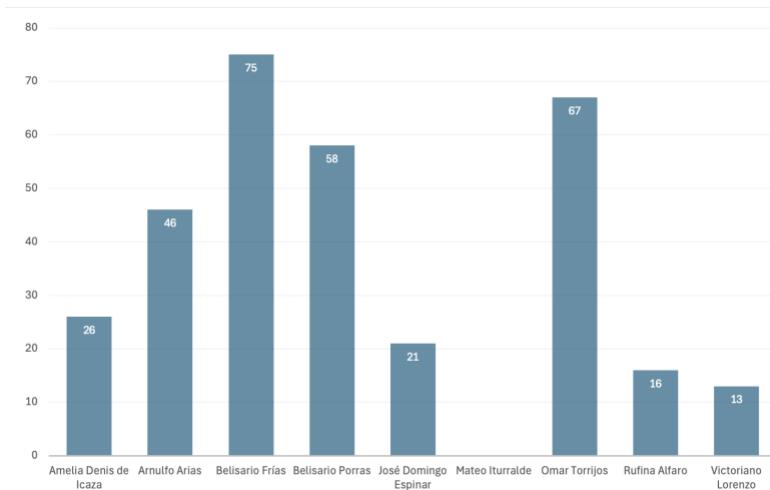


*Nota.* Elaboración propia en Excel (2025).

En la Figura B28 se señalan la distribución del conteo por corregimiento obtenido de los datos registrados de la ficha de observación.

**Figura B28**

*Cantidad de Acumulaciones Detectadas por Corregimientos*



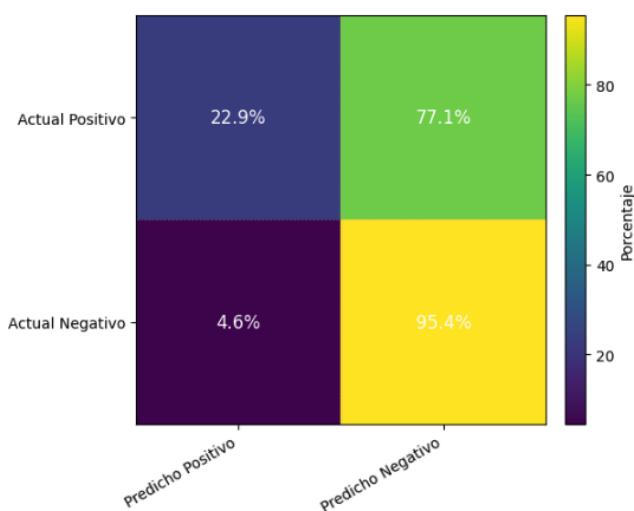
*Nota.* Elaboración propia en Excel (2025).

## Medir el Rendimiento General del Modelo: Fase de Evaluación de Generalización

A continuación, en la Figura B29, se presenta la matriz de confusión calculada con los valores reales registrados en la ficha de observación y las predicciones del modelo de visión computacional.

**Figura B29**

*Matriz de Confusión Imágenes Enero 2024*



*Nota.* Captura de pantalla de Jupyter Notebook local.

Seguidamente en la Figura B30 se muestran las métricas estándar aplicadas a los resultados de las predicciones del modelo de visión computacional.

**Figura B30**

*Resultados de Métricas del Modelo*

Métrica	Valor	Interpretación
0 Accuracy (precisión general)	86%	Proporción de aciertos totales
1 Recall (TPR / Sensibilidad)	23%	De los positivos reales
2 Precision (PPV) [opcional]	42%	De los predichos positivos
3 Specificity (TNR) [opcional]	95%	De los negativos reales

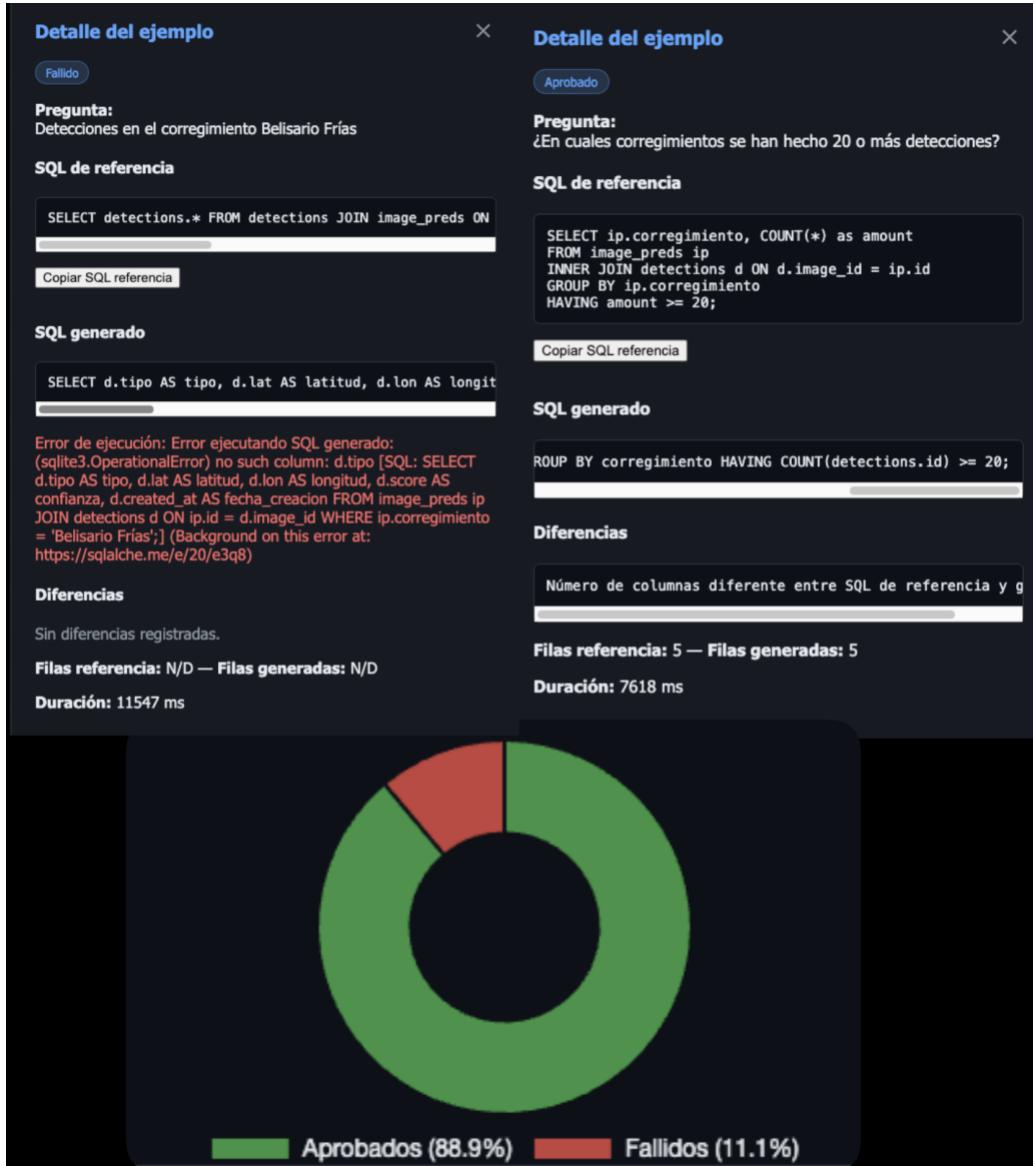
*Nota.* Captura de pantalla de Jupyter Notebook local.

## Evaluar la Fidelidad Explicativa-Semántica: Fase de Evaluación Simulada

En las Figuras B31 y B32 se muestran los resultados y dos ejemplos de los test automáticos ejecutados en el prototipo.

**Figura B31**

*Resultados de Execution Accuracy (EX)*



*Nota.* Captura de pantalla del prototipo (2025).

## Figura B32

### Resultados de Visual-Textual Similarity (VTS)

mean vtS: 23.016 median: 23.227 p75: 23.527 p90: 23.758 N=45 passed: 0 texto: llm LLM: gpt-4o prompt: default Status: finished

**Detalle del ítem**  
pred\_42.jpg  
Corregimiento: Victoriano Lorenzo



**Detalle del ítem**  
pred\_1733.jpg  
Corregimiento: Belisario Frías



**Texto generado (LLM)**  
En el corregimiento de Victoriano Lorenzo, se observa un foco puntual de basura dispersa en puntos aislados.

**Texto de referencia (gold)**  
En el corregimiento de Victoriano Lorenzo, sector resaltado muestra un foco puntual de basura disperso en puntos aislados.

**Métrica vtS**  
Estado: Falló (< 24.2)  
vtS: 22.5780  
Duración: 3117 ms  
Similitud texto: 72.7%  
Motivo: vtS por debajo del umbral configurado  
Fuente: LLM  
Chequeos lógicos: coherente con etiquetas  
Modelo LLM: gpt-4o  
Prompt: default  
Generado: 2025-09-28T05:28:39.067763  
Tokens: 683

**Texto generado (LLM)**  
El área evaluada está limpia; no se advierten bolsas ni botellas.

**Texto de referencia (gold)**  
No hay detecciones de residuos; se evalúa la imagen completa (total: 0).

**Métrica vtS**  
Estado: Falló (< 24.2)  
vtS: 20.7446  
Duración: 73 ms  
Similitud texto: 0.0%  
Motivo: vtS por debajo del umbral configurado  
Fuente: LLM  
Chequeos lógicos: coherente con etiquetas  
Modelo LLM: gpt-4o  
Prompt: default  
Generado: 2025-09-28T05:29:25.777374  
Tokens: 565

Nota. *Captura de pantalla del prototipo (2025).*

## **CONCLUSIONES**

Este estudio tiene como objetivo general implementar el prototipo con visión computacional explicable y semántica para la detección de acumulación de residuos en las calles de San Miguelito, Ciudad de Panamá. Además, planteamos que, para operacionalizar ese propósito se distribuyen las tareas específicas en fases de un proyecto de inteligencia artificial explicable y complementándose con los objetivos específicos de esta investigación, que son: construir el prototipo y contabilizar la acumulación de residuos en la fase de diseño y entrenamiento; medir el rendimiento general del modelo; y evaluar la fidelidad explicativa semántica. En conjunto, estas fases y objetivos medibles ponen a prueba, con evidencia empírica, la hipótesis formulada en esta investigación, que sostiene que la implementación del prototipo de visión computacional explicable y semántica permitirá detectar de forma fiable la acumulación de residuos en las calles. En esa misma línea, los hallazgos y análisis se organizan por fases, derivando en conclusiones generales.

### **Construir el Prototipo: Fase de Diseño y Entrenamiento**

La selección de la herramienta LandingLens, basada en el antecedente de Miller (2025), fue muy acertada, teniendo estas características como principales: ahorra tiempo en el etiquetado de imágenes por tener una interfaz espaciosa al buscar en la imagen; asimismo optimiza los entrenamientos por tener una Graphics Processing Unit (GPU) incluida en la versión gratuita de la plataforma; la curva de aprendizaje en el uso de la herramienta es corta gracias a la interfaz intuitiva confeccionada; contiene métricas estándar diversas, cada una con gráficos entendibles para el usuario no técnico; permite personalizar los parámetros del entrenamiento y usar diferentes tipos de augmentación de datos.

Además, el uso por parte de la plataforma LandingLens del modelo en visión computacional Representative Points (REPPONTS) del framework MMDetection

puede ser mejorado como base de pruebas en un entorno investigativo separado con mayores ejemplos de datos, debido a que es un framework de código abierto, como el utilizado en el antecedente Santamaría et al. (2022), que fue TensorFlow. Cabe señalar que la cantidad de créditos en el uso de la herramienta en la versión gratuita es baja; con los tres entrenamientos en esta fase del estudio, llegaron a consumirse de forma rápida, limitando técnicamente la mejora del rendimiento para el modelo que se usará en el prototipo. En cuanto a los resultados de esta fase de investigación, se ejecutaron tres entrenamientos, representados en la Figura B25. En el primer entrenamiento, etiquetado en la Figura B25 como “Entrenamiento 1”, se generó una precisión alta con 92%, y con sensibilidad (recall) relativamente baja con 85.5%. Es decir, el modelo es bueno detectando negativos y se le escapan 21 positivos.

En el segundo entrenamiento, etiquetado como “Entrenamiento 2” en la Figura B25, se agregan ejemplos positivos siguiendo el antecedente de Torres y Frernali (2023) en el uso del conjunto de datos público AerialWaste para mejorar la sensibilidad; luego se ejecuta el proceso con resultados de precisión similares al “Entrenamiento 1”, pero con caída significativa en la sensibilidad con 84.1% y precisión general con 67% en el conjunto “Test”, producido por el uso de los ejemplos del conjunto de datos AerialWaste. La causa detectada de esta caída fue por las diferencias geográficas del entorno urbano de las imágenes públicas AerialWaste con las de San Miguelito, Ciudad de Panamá. En el tercer y último entrenamiento, etiquetado con “Entrenamiento 3” en la Figura B25, se alcanza casi el límite de créditos de la plataforma LandingLens en la investigación.

En consecuencia, en este estudio se crea una secuencia de augmentación de datos en Python, cuyos parámetros son presentados en el Anexo B, Figura B8, para optimizar créditos. Después de realizar la ejecución, los resultados aumentaron significativamente en cuanto a precisión general en el conjunto “test” con 92%, manteniendo cifras similares en precisión y sensibilidad, debido a no tener más imágenes disponibles con identificación positiva dentro de las delimitaciones geográficas del estudio. Además, como referencia, en el estudio se

incluyen las métricas Mean Average Precision (MAP) de cada entrenamiento, expuestas en la Figura B26, que señalan que la precisión para localizar objetos dentro de “Entrenamiento 1” inició con 0.7 de media de localización; luego, en el “Entrenamiento 2” bajó a 0.6 de media aproximadamente, y el “Entrenamiento 3” subió a 0.8. El prototipo alcanzó métricas suficientes para cumplir con el objetivo en la fase de entrenamiento.

### **Contabilizar la Acumulación de Residuos: Fase de Diseño y Entrenamiento**

En esta fase de la investigación, los gráficos de la Figura B27 y B28 muestran el conteo y la distribución espacial de las acumulaciones detectadas en el distrito de San Miguelito, Ciudad de Panamá, por medio de la inspección sistemática de las imágenes satelitales con el instrumento ficha de observación. Estas figuras señalan a los corregimientos de Belisario Frías, con 75 detecciones (21%), y Omar Torrijos, con 67 detecciones (21%), como los corregimientos con más acumulaciones. Es preciso señalar que cada distribución espacial detectada confirma la correlación ya expuesta por Bernal (2024) entre los vectores de contaminación y la gestión de residuos, con las consecuencias de la propagación de mosquitos, como el dengue. Además, da visibilidad de la magnitud del problema no solo en San Miguelito, sino en las zonas circundantes, en consonancia con lo expuesto por Pitty y Migdaris (2021).

En consecuencia, la alerta emitida por el Ministerio de Salud en enero 2024 confirma lo expresado por los datos de la investigación. Estos datos cercanos a la realidad son la línea base necesaria para evaluar de forma robusta el rendimiento del prototipo en consonancia con la realidad del distrito de San Miguelito, Ciudad de Panamá y cumplen con el objetivo específico de la contabilización de los residuos.

## **Medir el Rendimiento General del Modelo: Fase de Evaluación de Generalización**

En esta etapa, la matriz de confusión que compara la realidad señalada de acumulación de residuos en la fase anterior con las predicciones del modelo, es representada en la Figura B29, que indica lo siguiente: verdadero positivo: muestra cuando el modelo predice positivo y la etiqueta actual es positiva, acierta un 22.9%; falsos positivos: señala que, cuando el modelo predice positivo y la etiqueta actual es negativa, se equivoca en un 4.6%; negativos verdaderos: cuando el modelo indica que es negativo y la etiqueta actual es negativa, acierta un 95.6%; falsos negativos: cuando el modelo predice negativo y lo real es positivo, se equivoca en un 77.1%. En cuanto a la Figura B30, las métricas de rendimiento derivadas de la matriz de confusión indican la precisión general del 86%, superando la seleccionada como umbral en el antecedente Cárdenas et al. (2024) con un 84%. Por lo tanto, el modelo demostró en la evaluación de generalización que el rendimiento es adecuado superando el umbral definido de 84% cumpliendo con el objetivo.

## **Evaluar la Fidelidad Explicativa-Semántica: Fase de Evaluación Simulada**

En esta fase del estudio, los resultados obtenidos por las métricas evaluativas de semántica Execution Accuracy (EX), explicada por Hong et al. (2024), y la explicabilidad Visual-Textual Similarity (VTS) de Li et al. (2024) indican lo siguiente: EX muestra en la Figura B31 que el Large Language Model (LLM) GPT-4o su fidelidad semántica a nivel consultas estructuradas a base de datos, es correcta el 88.9% de las veces y falla solo un 11%, considerando que el prototipo es evaluado con consultas básicas para limitar el uso de tokens generados, evitando costos adicionales.

Asimismo, en la Figura B31 se presentan dos ejemplos que determinan específicamente qué mide EX: el primero muestra un error de unos de los ejemplos incluidos con errores intencionales en la consulta generada por el modelo, que

indica que la columna “d.tipo” no existe, demostrando que, con el uso de la métrica de Hong et al. (2024), se detectan esos casos donde el LLM GPT-4o infiere erróneamente; el segundo ejemplo indica que es aprobado porque la salida de la consulta de base de datos es igual en 5 filas y columnas, demostrando la esencia de la métrica de Hong et al. (2024), que es validar que el resultado de referencia sea el mismo que la salida generada por el modelo.

En cuanto a la métrica VTS que se presenta en la Figura B32, la ejecución de pruebas automatizadas en el prototipo arroja que las explicaciones generadas son adecuadas de forma general, representadas cuantitativamente con una media de 23.01, aunque este valor es menor a la recomendación de Li et al. (2024), de utilizar la media de las últimas pruebas cuyo valor en esta investigación es de 24.2 de la media de 500 pruebas anteriores. Estos resultados son coherentes con lo expuesto como aporte del antecedente de Pitty y Migdaris (2021), al describir el alcance del impacto por la contaminación del vertedero es por áreas cercanas y no por lugares específicos, es decir, al ampliar la zona de la descripción del LLM, el indicador de la media VTS disminuye sin perder totalmente la fidelidad explicativa. Por lo tanto, se determina que el prototipo logra con el objetivo específico en explicar de forma fiel y multimodal sus decisiones con valores aceptables como media 23.01.

En conclusión, la investigación cumplió con el objetivo general de implementar un prototipo de visión computacional explicable-semántico para detectar acumulaciones de residuos en las calles de San Miguelito, Ciudad de Panamá, cumpliendo con cada objetivo específico al contabilizar la acumulación de residuos y construir el prototipo en la fase de diseño y entrenamiento; al medir el rendimiento general del modelo en la fase de evaluación de generalización; y por último, al evaluar la fidelidad explicativa-semántica en la fase de evaluación simulada con una media . Cada fase valida la hipótesis de que el prototipo permitirá detectar de forma fiable la acumulación de residuos en San Miguelito, Ciudad de Panamá. Por lo tanto, los aportes de esta investigación son los siguientes: teóricos al aplicar métricas de fidelidad en inteligencia artificial explicable a un caso

ambiental; en lo práctico, dando una herramienta comprensible y transparente para la gestión de residuos; y en lo metodológico, con la integración de visión computacional con explicabilidad textual, visual y semántica en consultas estructuradas a base de datos.

## LIMITACIONES

Este estudio está expuesto a algunas limitaciones como las siguientes: el conjunto de datos está limitado por la dependencia de la disponibilidad de imágenes satelitales hasta enero de 2024 por parte de Google Earth Pro. Además, la validación se realizó únicamente en entornos simulados y no se pudo realizar una verificación de campo de los sitios de acumulaciones detectadas, lo que impide la prueba empírica de las detecciones del prototipo; a nivel técnico, existen restricciones en la cantidad de créditos disponibles a través de la plataforma LandingLens y en la capacidad de procesamiento de la interfaz de programación de aplicaciones del modelo Generative Pre-trained Transformer 4 omni (GPT-4o), lo que nos impidió realizar experimentos con repositorios de imágenes más grandes.

También, las restricciones metodológicas incluyeron la limitación geográfica de la muestra y población al distrito de San Miguelito, Ciudad de Panamá; por lo tanto, los resultados no pueden generalizarse fácilmente a otros distritos ni áreas urbanas en la Ciudad de Panamá. Finalmente, en este estudio se utilizó una implementación experimental de prueba de concepto, es posible que la fidelidad del prototipo pueda cambiar en función de varios factores, como la resolución de la imagen, o las condiciones atmosféricas. Sin embargo, estas limitaciones no afectan la validez técnica de los resultados, pero informan a las investigaciones a futuro para validar más a fondo.

## **RECOMENDACIONES**

Las recomendaciones para líneas futuras de investigación son: incorporar imágenes satelitales de períodos más actualizados usando herramientas como ArcGIS; validar delimitaciones administrativas de los corregimientos con fuentes gubernamentales; comprobar las detecciones con inspecciones de campo; utilizar imágenes georreferenciadas como Georeferenced Tagged Image File Format (GEOTIFF); ampliar a otros distritos o ciudades el conjunto de imágenes; explorar nuevas métricas de explicabilidad y semántica; comparar explicaciones textuales y visuales con otros modelos de lenguaje de gran escala.

Además, se sugiere superar las restricciones técnicas, como la limitación de créditos disponibles en la plataforma LandingLens y los límites de cuota de procesamiento de la interfaz de programación de aplicaciones (API) del modelo Generative Pre-trained Transformer 4 omni (GPT-4o), estas acciones permitirían mejorar la experimentación con repositorios de imágenes más grandes. Por último, para asegurar la utilidad práctica, se recomienda trabajar en la integración con la interfaz de aplicaciones de la plataforma del operador responsable de la recolección de residuos en el distrito de San Miguelito, Ciudad de Panamá.

## REFERENCIAS

- Alarcón Montero, P. A., Acosta Acevedo, S., Correal, M., Piamonte, C., Rihm, A., Breukers, L., Durón, L., González, G., Hernández López, C., Sagasti, C., y Rojas Gutiérrez, A. (2023). *Evaluación regional de flujo de materiales: Residuos sólidos municipales para América Latina y el Caribe (EVAL 2023)* (B. Fallik y C. M. Pasquetti, Eds.). <https://doi.org/10.18235/0004841>
- Astobiza, A. M. (2025). The role of LLMs in theory building. *Social Sciences & Humanities Open*, 11, 101617. <https://doi.org/10.1016/j.ssho.2025.101617>
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Bernal Mosquera, D. (2024). *Situación del manejo de residuos municipales en el Área Metropolitana de la Ciudad de Panamá y su influencia en el desarrollo de vectores* [Tesis de maestría, Universidad Miguel Hernández]. <https://hdl.handle.net/11000/31717>
- Bi, Y., Xue, B., Mesejo, P., Cagnoni, S., & Zhang, M. (2022). A Survey on Evolutionary Computation for Computer Vision and Image Analysis: Past, Present, and Future Trends. *ArXiv*. <https://doi.org/10.1109/TEVC.2022.3220747>
- C de Comunicación. (2023, 19 de septiembre). *ABB reducirá un 80 % el entrenamiento de visión artificial de sus robots y acortará tiempos de implementación en sectores como el logístico*. *Revista Logística*. <https://logistica.cdecomunicacion.es/automatizacion-y-robotica/155553/abb-reducira-un-80-el-entrenamiento-de-vision-artificial-de-sus-robots-y-acortara-tiempos-de-implementacion-en-sectores-como-el-logistico>
- Cañete Marzo, C. (2024). *Detección de basurales a cielo abierto en nuestra región mediante el uso de inteligencia artificial: Generación de los datos locales* [Resumen de tesis de licenciatura, Universidad Nacional del Litoral]. <https://hdl.handle.net/11185/8476>

- Cárdenas, L., Pipino, A., Donadio, M. T., y Picota, P. (2024). Modelo de un sistema de alerta basado en visión artificial para la prevención de aglomeraciones en el transporte público. *Visión Antataura*, 8(2), 93–107.  
<https://doi.org/10.48204/j.vian.v8n2.a6571>
- DataScientest. (2021, 14 de diciembre). *¿Qué es el método Grad-CAM?*  
<https://datascientest.com/es/que-es-el-metodo-grad-cam>
- DataScientest. (2024a, 29 de julio). *Transformer preentrenado generativo (GPT): ¿qué es?* <https://datascientest.com/es/transformer-preentrenado-generado-gpt-que-es>
- DataScientest. (2024b, 29 de febrero). *Matriz de confusión.*  
<https://datascientest.com/es/matriz-de-confusion>
- Domínguez, M. O. (2025). Redes neuronales artificiales. *Ingenio y Conciencia – Boletín Científico de la Escuela Superior Ciudad Sahagún*, 12(23), 38–44.
- Esri. (s. f.-a). *GeoJSON*. ArcGIS Enterprise.  
<https://enterprise.arcgis.com/es/portal/latest/use/geojson.htm>
- Esri. (s. f.-b). *Essential KML vocabulary*. ArcGIS Documentation.  
<https://doc.arcgis.com/es/allsource/latest/data/essential-kml-vocabulary>
- Expansión. (2025, 20 de febrero). *Large World Models: La nueva frontera de la inteligencia artificial*. <https://www.expansion.com/expansion-empleo/emprendedores/2025/02/20/67b75ddce5fdea54348b4574.html>
- Fundación Bunge y Born. (2021). *Detección de basurales a cielo abierto*.  
<https://www.fundacionbyb.org/basurales>
- Goellner, S., Tropmann-Frick, M., y Brumen, B. (2024). Inteligencia artificial responsable: Una revisión estructurada de la literatura [Preprint]. arXiv.  
<https://doi.org/10.48550/arXiv.2403.06910>
- Gonzalez Mesones, E. (2023, 13 de septiembre). *Explicabilidad de modelos de ML: LIME*. Medium. <https://medium.com/latinixinai/explicabilidad-de-modelos-de-ml-lime-f9d0dceb5154>
- Google Developers. (s.f.-a). *Tutorial de KML*. Google Developers.  
[https://developers.google.com/kml/documentation/kml\\_tut?hl=es-419](https://developers.google.com/kml/documentation/kml_tut?hl=es-419)

Google Developers. (s.f.-b). *Exactitud, precisión y recuperación*. Google Developers.

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=es-419>

Google Developers.(s.f.-c). *Glosario de aprendizaje automático: Conceptos básicos del AA*. Google for Developers. <https://developers.google.com/machine-learning/glossary?hl=es-419>

Greenpeace Colombia. (2023, 30 de mayo). *Llenamos al mundo de basura, ¿y ahora qué?* Greenpeace.

<https://www.greenpeace.org/colombia/blog/issues/contaminacion/llenamos-almundo-de-basura-y-ahora-que/>

Hong, Z., Yuan, Z., Zhang, Q., Chen, H., Dong, J., Huang, F., & Huang, X. (2024). *Next-Generation Database Interfaces: A Survey of LLM-based Text-to-SQL*. ArXiv. <https://arxiv.org/abs/2406.08426>

IES Cerro del Viento. (2023). *Google Maps y Google Earth*.

<https://iescerrodelviento.es/posts/google-maps-y-google-earth/498>

Isaza, I. I., Bethancourt, H. I., y Aguirre, A. A. (2025). Análisis del área y calidad del suelo del relleno sanitario del Jobo de Antón en la provincia de Coclé mediante el uso de la teledetección. *Revista Científica Guacamaya*, 9(2), 32–45.

Jonker, A. (2023, 20 de noviembre). *Sistemas de información geográfica (SIG)*. IBM Think. <https://www.ibm.com/mx-es/think/topics/geographic-information-system>

Karn, A. (2021). Artificial intelligence in computer vision. *International Journal of Engineering Applied Sciences and Technology*, 6(1), 249–254.

Kotz, G., Salcedo, P., y Fuentes, K. (2024). Análisis léxico de textos generados por modelos de lenguaje: Reflejo de sus modelos de mundo. *Lengua y Sociedad*, 23(2), 895–910. <https://doi.org/10.15381/lengsoc.v23i2.28336>

República de Panamá. (2021, 30 de diciembre). *Ley N.º 276 de 2021 que regula la gestión integral de residuos sólidos en la República de Panamá. Gaceta Oficial Digital* (29,445-E).

[https://www.gacetaoficial.gob.pa/pdfTemp/29445\\_E/GacetaNo\\_29445e\\_20211230.pdf](https://www.gacetaoficial.gob.pa/pdfTemp/29445_E/GacetaNo_29445e_20211230.pdf)

- Li, K., Vosselman, G., & Yang, M. Y. (2024). *Explainable visual question answering with multimodal rationales*. arXiv. <https://arxiv.org/html/2402.03896v2>
- Lu, P., Mishra, S., Xia, T., Qiu, L., Chang, K., Zhu, S., Tafjord, O., Clark, P., & Kalyan, A. (2022). Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering. ArXiv. <https://arxiv.org/abs/2209.09513>
- McGregor, M. (2024, 28 de septiembre). *GPT-4: Guía completa*. Semrush Blog. <https://es.semrush.com/blog/gpt-4-guia-completa/>
- Miller, M. (2025). *Exploring the potential of no-code artificial intelligence in identifying periapical radiolucent lesions* [Tesis de maestría, Marquette University]. [https://epublications.marquette.edu/theses\\_open/849/](https://epublications.marquette.edu/theses_open/849/)
- Ministerio de Salud de la República de Panamá [MINSA]. (2024, 26 de enero). *Declaran alerta sanitaria en San Miguelito por deficiencia en la recolección de la basura*. <https://www.minsa.gob.pa/noticia/declaran-alerta-sanitaria-en-sanmiguelito-por-deficiencia-en-la-recoleccion-de-la-basura>
- Ministerio de Salud de la República de Panamá [MINSA]. (2025, 16 de marzo). *Acciones de la población claves para eliminar criaderos y evitar el dengue*. <https://www.minsa.gob.pa/noticia/acciones-de-la-poblacion-claves-para-eliminarcriaderos-y-evitar-el-dengue>
- Mohammadjafari, A., Maida, A. S., & Gottumukkala, R. (2024). From Natural Language to SQL: Review of LLM-based Text-to-SQL Systems. ArXiv. <https://arxiv.org/abs/2410.01066>
- Moin, T. S. (2023). *Overview of computer vision*. Universidad de Dhaka. <https://doi.org/10.13140/RG.2.2.13989.68327>
- Molnar, C. (2019). *Interpretable machine learning*. <https://christophm.github.io/interpretable-ml-book/>
- Morales, A. (2018, 8 de enero). *Overpass turbo: Minería de datos para OpenStreetMap*. MappingGIS. <https://mappinggis.com/2018/01/overpass-turbo-mineria-de-datos-para-openstreetmap/>
- Navarrete Salazar, M. G. (2024). *Diseño e implementación de un sistema de segmentación de componentes electrónicos en tarjetas electrónicas mediante*

*visión artificial* [Tesis de grado, Universidad Politécnica Salesiana].

<https://dspace.ups.edu.ec/handle/123456789/27729>

Neuroons. (2021, 22 de enero). *¿Qué son entrenamiento e inferencia en inteligencia artificial?* <https://neuroons.com/es/que-son-entrenamiento-e-inferencia-en-inteligencia-artificial/>

OSGeo. (s. f.). *Quantum GIS (QGIS) overview*. OSGeo-Live 9.0.

[https://live.osgeo.org/archive/9.0/es/overview/qgis\\_overview.html](https://live.osgeo.org/archive/9.0/es/overview/qgis_overview.html)

Páez, A. (2025). *Transparencia, explicabilidad y confianza en los sistemas de aprendizaje automático*. ResearchGate.  
[https://www.researchgate.net/publication/385705473\\_Transparencia\\_explicabilidad\\_y\\_confianza\\_en\\_los\\_sistemas\\_de\\_aprendizajeAutomatico](https://www.researchgate.net/publication/385705473_Transparencia_explicabilidad_y_confianza_en_los_sistemas_de_aprendizajeAutomatico)

Parra Méndez, P. (2023). *Detección de lesiones hepáticas en ecografías intraoperatorias mediante Deep Learning* [Trabajo Fin de Máster, Universidad Loyola]. Repositorio Institucional Universidad Loyola.  
<https://hdl.handle.net/20.500.12412/6736>

Palacio, S., Lucieri, A., Munir, M., Hees, J., Ahmed, S., & Dengel, A. (2021). XAI Handbook: Towards a Unified Framework for Explainable AI. *ArXiv*.  
<https://arxiv.org/abs/2105.06677>

Pascasio, J. J., Mela, R. H., Vélez, M. L., y Rangel, J. C. (2022). Implementación de redes neuronales para la clasificación de desechos dentro de un cesto inteligente. *Centros: Revista Científica Universitaria*, 11(1), 229–245.

Pineda Pertuz, C. M. (2022). *Aprendizaje automático y profundo en Python*. Ra-Ma Editorial.

<https://books.google.com.pa/books?id=mgNcEAAAQBAJ&lpg=PA1&hl=es&pg=PA1#v=onepage&q&f=false>

Pitty, M., y Migdaris, Y. (2021). *Evaluación de impactos y problemáticas existentes en las comunidades aledañas al relleno sanitario Cerro Patacón, Ciudad de Panamá* [Informe no publicado]. [Tesis de grado, Universidad de Panamá].  
<https://up-rid.up.ac.pa/6563/>

Programa de las Naciones Unidas para el Medio Ambiente. (2020, 7 de septiembre). *Cerrar basurales para limpiar el aire en América Latina y el Caribe*.

<https://www.unep.org/es/noticias-y-reportajes/reportajes/cerrar-basurales-para-limpiar-el-aire-en-america-latina-y-el>

Reche, C., Tobias, A., & Viana, M. (2022). *Vehicular Traffic in Urban Areas: Health Burden and Influence of Sustainable Urban Planning and Mobility*. *Atmosphere*, 13(4), 598. <https://doi.org/10.3390/atmos13040598>

Rommel, F. (s. f.). *SQLite: La base de datos embebida*. Software Gurú.

<https://sg.com.mx/revista/17/sqlite-la-base-datos-embebida>

Tarrillo Saldaña, O., Mejía Huamán, J., Dávila Mego, J. S., Pintado Castillo, C. A., Tapia Idrogo, C. E., Chilón Camacho, W. M., & Vélez Escobar, S. B. (2024). *Metodología de la investigación: Una mirada global. Ejemplos prácticos*. CID – Centro de Investigación y Desarrollo. [https://doi.org/10.37811/cli\\_w1078](https://doi.org/10.37811/cli_w1078)

Santamaría, K., Solé, M., y Rodríguez, K. (2022). Reconocimiento de tipos de serpientes en Panamá por medio de visión artificial. *Revista de Iniciación Científica*, 8(1), 65–70.

Santos Jr., C., & Dorneles, C. F. (2022). *Enriching georeferenced environmental data using web data extraction to contribute to degraded area impact analysis*. En C.-P. Rückemann & Y. Doytsher (Eds.), *Proceedings of the 14th International Conference on Advanced Geographic Information Systems, Applications and Services (GEOProcessing 2022)* (pp. 31–35). Porto, Portugal: IARIA.

Sarkar, A., Idris, M. Y. I., & Yu, Z. (2025). *Reasoning in Computer Vision: Taxonomy, Models, Tasks, and Methodologies* (arXiv:2508.10523). arXiv. <https://doi.org/10.48550/arXiv.2508.10523>

Schwalbe, G., Finzel, B. (2024) . A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. *Data Min Knowl Disc* 38, 3043–3101 (2024). <https://doi.org/10.1007/s10618-022-00867-8>

Secretaría Nacional de Ciencia, Tecnología e Innovación [SENACYT]. (2020, 22 de noviembre). *Nota de prensa – Ceremonia*. SENACYT. <https://www.senacyt.gob.pa/wp-content/uploads/2020/11/Nota-de-prensaCeremonia.pdf>

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., y Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based

- localization. En *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626). <https://doi.org/10.1109/ICCV.2017.74>
- Sirkin, J. (2021, 22 de abril). SQL (*Structured Query Language o Lenguaje de consultas estructuradas*). Computer Weekly en Español.  
<https://www.computerweekly.com/es/definicion/SQL-Structured-Query-Language-o-Lenguaje-de-consultas-estructuradas>
- Striker, C., y Bergmann, D. (2025, 28 de marzo). *Modelo de transformador*. IBM Think.  
<https://www.ibm.com/es-es/think/topics/transformer-model>
- Torregrosa Domínguez, Á. (2023). *Detección orientada de objetos en imágenes satelitales de alta resolución* [Trabajo Fin de Máster, Universidad Politécnica de Madrid]. Archivo Digital UPM.  
[https://oa.upm.es/75859/1/TFM\\_ANGEL\\_TORREGROSA\\_DOMINGUEZ.pdf](https://oa.upm.es/75859/1/TFM_ANGEL_TORREGROSA_DOMINGUEZ.pdf)
- Torres, R. N., & Fraternali, P. (2023). AerialWaste dataset for landfill discovery in aerial and satellite images. *Scientific Data*, 10(1), 1-14. <https://doi.org/10.1038/s41597-023-01976-9>
- Tuychiev, B. (2025, 9 de abril). *OpenAI responses API*. DataCamp.  
<https://www.datacamp.com/es/tutorial/openai-responses-api>
- Wan, A. A. (2024, 14 de marzo). *Introducción a los valores SHAP: Interpretabilidad de machine learning*. DataCamp. <https://www.datacamp.com/es/tutorial/introduction-to-shap-values-machine-learning-interpretability>
- Zeron, A. (2023). Inteligencia artificial, inteligencia emocional e inteligencia intestinal. *Revista ADM – Órgano Oficial de la Asociación Dental Mexicana*, 80(3), 126–128.

# ANEXO A

## Tablas de la Sección Marco Teórico

**Tabla A1**

*Resumen de Antecedentes y su Aporte*

Antecedente	Aporte
<b>Acumulación de Residuos en las Calles</b>	
Análisis del Área y Calidad del Suelo del Relleno Sanitario del Jobo de Antón en la Provincia de Coclé	Se adopta el historial de Google Earth Pro para formar conjuntos de entrenamiento de abril 2023 y validación de enero 2024.
Mediante el uso de la Teledetección (Isaza et al., 2025).	
Situación del Manejo de Residuos Municipales en el Área Metropolitana de la Ciudad de Panamá y su Influencia en el Desarrollo de Vectores (Bernal, 2024).	La relación gestión de residuos–vectores respalda la operacionalización y medición cuantitativa de variables.
Enriquecer Bases de Datos con Información Ambiental Georreferenciada Mediante la Extracción de Datos Procedentes de Archivos KML Disponibles en la Web (Santos y Dorneles, 2022).	Se emplean archivos KML para georreferenciación; se delimitan celdas 150 m y se sistematiza la captura de imágenes.
Vehicular Traffic in Urban Areas: Health Burden and Influence of Sustainable Urban Planning and Mobility (Reche et al., 2022)	Se replica la obtención de límites administrativos con Overpass Turbo para delimitar corregimientos en San Miguelito.

<b>Antecedente</b>	<b>Aporte</b>
Evaluación de Impactos y Problemáticas Existentes en las Comunidades Aledañas al Relleno Sanitario Cerro Patacón, Ciudad de Panamá, Panamá (Pitty y Migdaris, 2021).	Se analiza por zonas cercanas a acumulaciones, segmentando en celdas 150 m con solapamiento definido.
<b>Antecedente</b>	<b>Aporte</b>
<b>Visión Computacional</b>	
Exploring the Potential of No-Code Artificial Intelligence in Identifying Periapical Radiolucent Lesions (Miller, 2025).	Uso de LandingLens para etiquetado y entrenamiento del modelo.
Modelo de un Sistema de Alerta Basado en Visión Artificial para la Prevención de Aglomeraciones en el Transporte Público (Cárdenas et al., 2024)	Fija la precisión general como métrica clave con umbral de aceptación del 84.04 %.
Reconocimiento de Tipos de Serpiente en Panamá por Medio de Visión Artificial (Santamaría et al., 2022)	Uso de frameworks abiertos MMDetection integrados en el prototipo.
<b>Antecedente</b>	<b>Aporte</b>
<b>Visión Computacional para Detectar Residuos</b>	
Detección de Basurales a Cielo Abierto en Nuestra Región Mediante el Uso de Inteligencia Artificial: Generación de los Datos Locales (Cañete, 2024)	Obtención de imágenes satelitales con Google Earth Pro y derivación de la ficha de observación.

<b>Antecedente</b>	<b>Aporte</b>
Aerial Waste Dataset for Landfill Discovery in Aerial and Satellite Images (Torres y Fraternali, 2023)	Incorpora una muestra de AerialWaste para ampliar cobertura y mejorar precisión.
Implementación de Redes Neuronales para la Clasificación de Desechos Dentro de un Cesto Inteligente (Pascasio et al., 2022)	Guarda en SQLite predicciones y coordenadas.
<b>Antecedente</b>	<b>Aporte</b>
<b>Visión Computacional Explicable</b>	
Explainable Visual Question Answering with Multimodal Rationales (Li et al., 2024).	Evalúa fidelidad explicativa con VTS mediante pruebas automatizadas.
Image-of-Thought Prompting for Visual Reasoning Refinement in Multimodal Large Language Models (Zhou et al., 2024).	Configuración del módulo semántico con GPT-4o con temperatura cercana a cero.
Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering (Lu et al., 2022)	Razonamiento encadenado para consultas SQL, ejecución y explicación narrativa.
GRAD-CAM: Visual Explanations from Deep Networks via Gradient-based Localization (Selvaraju et al., 2017)	Base post-hoc para visualización de regiones relevantes en CNN que guía la explicabilidad del prototipo.

*Nota.* Elaboración propia a partir de diversas fuentes.

**Tabla A2***Criterios para Aprendizaje Automático*

<b>Criterio</b>	<b>Concepto</b>
Generación de explicación	Intrínsecamente interpretable: el propio modelo es transparente.
	Explicación post-hoc: explicación después de entrenar un modelo oculto.
Tipo de explicación	Global: comportamiento en todos los casos.
	Local: comportamiento en casos específicos.
Audiencia o propósito	Técnicas, no técnicas o de propósito como depuración, cumplimiento legal, etc.
Método empleado	Perturbaciones de la entrada, en gradientes, sustitutos, ejemplos contrafactuales.

*Nota.* Elaboración propia a partir de Barredo et al. (2020).

**Tabla A3***Resumen de Concepto y su Aporte*

<b>Autor - Concepto</b>	<b>Aporte resumido</b>
Astobiza (2025) – Construcción de teorías con LLM	Apoya la formulación de hipótesis sobre patrones de residuos.
Astobiza (2025) – Limitaciones de LLM	El prototipo depende de validación humana y técnica por límites de los modelos de lenguajes de gran escala.
Astobiza (2025) – Modelos de Lenguaje de Gran Escala (LLM)	Define GPT-4 como motor de consulta en lenguaje natural en el prototipo.

Astobiza (2025) – Representación semántica en LLM	Anclar las respuestas a datos SQL para evitar respuestas abstractas.
Hong et al. (2024) – Componentes clave del Text-to-SQL	Aplicar comprensión de lenguaje, esquema de base de datos y generación de SQL.
Hong et al. (2024) – Métricas de evaluación del Text-to-SQL	Métrica EX, para evaluar la precisión a nivel SQL.
Hong et al. (2024) – Objetivo del Text-to-SQL	Facilita consultas de usuarios no expertos con generación automática de SQL.
Hong et al. (2024) – Text-to-SQL con LLM	El prototipo convierte consultas en lenguaje natural a SQL ejecutable.
Navarrete (2024) – Tipos de visión computacional	Se aplica la detección de objetos, alineada con la tarea del prototipo.
Bi et al. (2022) – Visión computacional	Principales funciones del prototipo que son extraer, analizar y comprender imágenes.
Barredo et al. (2020) – Conceptos fundamentales de inteligencia artificial explicable	Se aplican interpretabilidad, explicabilidad y transparencia en las salidas del prototipo.
Barredo et al. (2020) – Desafíos de la inteligencia artificial explicable	Se atienden fidelidad de explicaciones y diversidad de audiencias con métricas y respuestas diferenciadas.
Barredo et al. (2020) – Fases en proyectos de Inteligencia Artificial Explicable (XAI)	Mayor acercamiento al objetivo de la investigación.
Barredo et al. (2020) – Inteligencia Artificial Explicable (XAI)	Aporta transparencia y comprensibilidad al modelo REPOINTS.
Barredo et al. (2020) – Inteligencia artificial responsable	Sustenta que el prototipo dé explicaciones claras y transparentes.

Barredo et al. (2020) – Oportunidades de la XAI	El prototipo implementa confianza, cumplimiento legal y mejor interacción humano–máquina.
Barredo et al. (2020) – Taxonomía de XAI	Orienta la generación de explicaciones post-hoc locales y globales adaptadas a la audiencia.
Ley N.º 276 (2021) – Definición de residuo	Aporta el marco legal para el uso del término residuo.
Ley N.º 276 (2021) – Responsabilidad de la Autoridad de Aseo	Base legal complementada con detección automática en el prototipo.

---

*Nota.* Elaboración propia a partir de diversas fuentes.

# ANEXO B

## Tablas y Figuras de la Sección Metodología

**Tabla B1**

*Operacionalización de Variables*

Variable	Definición conceptual	Definición operac.	Dimensión	Indicadores	Regla de decisión
	Sistema basado en modelos de detección de objetos y explicabilidad		Rendimiento del modelo	Precisión general (%)	Basada en Cárdenas et al. (2024): umbral de aceptación $\geq 84.04\%$ .
Independiente : Prototipo con visión computacional explicable-semántico	semántica que identifica acumulacione s de residuos en imágenes satelitales, justificando la decisión con lenguaje natural y consultas SQL.	Medida por el informe de evaluació n del prototipo.	Fidelidad explicativa	Execution Accuracy (%)	Basado en Hong et al. (2024): aprobado sólo si el SQL generado coincide con el esperado en las mismas columnas y filas.

Variable	Definición conceptual	Definición operac.	Dimensión	Indicadores	Regla de decisión
Dependiente: Acumulación de residuos en las calles	Presencia de desechos visibles en espacios públicos (calles, aceras, lotes baldíos).	Medida con ficha de obser. derivada de Cañete (2024) aplicada a imágenes satelitales de enero 2024.	Fidelidad explicativa  Visual–Textual  Similarity (Promedio)	Cantidad de residuos por imagen.	Basado en Li et al. (2024): fidelidad explicativa multimodal, aprobada si se mantiene cercana al umbral de referencia de pruebas previas.

Variable	Definición conceptual	Definición operac.	Dimensión	Indicadores	Regla de decisión
		Identificación de la unidad de análisis	Código único de la imagen satelital (nombre de archivo).		Permite trazabilidad y evita duplicidad en el registro.

*Nota.* Elaboración propia.

**Tabla B2**

*Teselas de la Población por Corregimiento*

Corregimiento	Cantidad de teselas
Amelia Denis de Icaza	171
Arnulfo Arias	151
Belisario Frías	218
Belisario Porras	181
José Domingo Espinar	320
Mateo Iturralde	47
Omar Torrijos	483
Rufina Alfaro	568
Victoriano Lorenzo	88
Total	2227

*Nota.* Elaboración propia.

**Tabla B3***Distribución Primer Entrenamiento*

Hiperparámetros		Augmentación	Distribución de las imágenes	
Épocas	40	Cambio Horizontal	Entrenamiento	443 (70%)
Tamaño del modelo	37 M.	Prob. 0.5	Validación	125 (20%)
			Prueba	64 (10%)
			Total	632

*Nota.* Elaboración propia a partir de los datos en LandingLens (2025).

**Tabla B4***Distribución Segundo Entrenamiento*

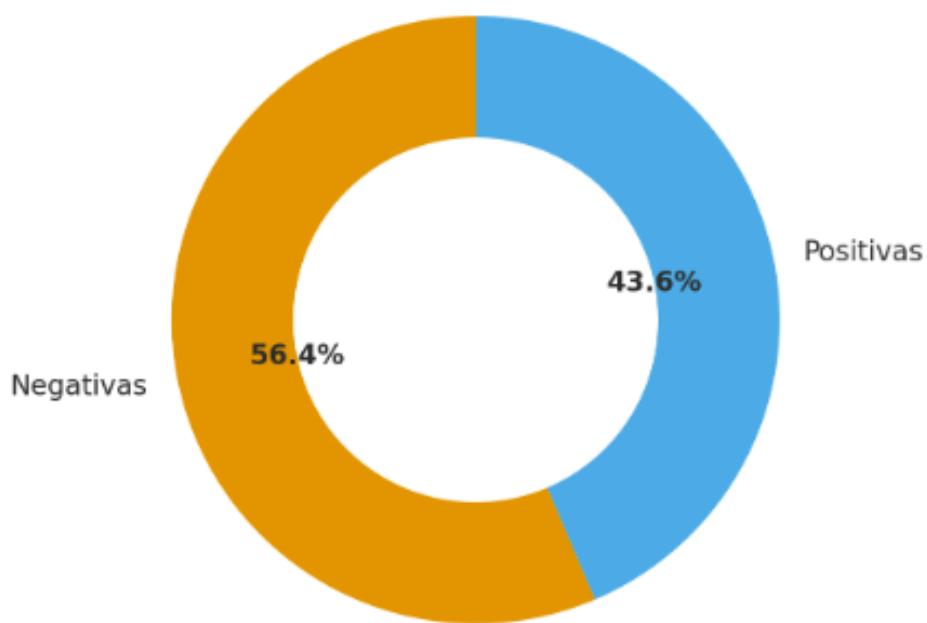
Hiperparámetros		Augmentación	Distribución de las imágenes	
Épocas	40		Entrenamiento	584 (70%)
Tamaño del modelo	37 M.	Cambio Horizontal	Validación	161 (20%)
		Prob. 0.5	Prueba	64 (10%)
			Total	789

*Nota.* Elaboración propia a partir de los datos en LandingLens (2025).

**Tabla B5***Distribución Tercer Entrenamiento*

Hiperparámetros	Augmentación	Distribución de las imágenes	
Épocas	40	Con augmentación	583 (70%)
Tamaño del modelo	37 M.	manual	168 (20%)
			Prueba
			83 (10%)
			789

*Nota.* Elaboración propia a partir de los datos LandingLens (2025).

**Figura B1***Distribución de Imágenes de Entrenamiento*

*Nota.* Elaboración propia a partir de los datos en LandingLens (2025)

**Figura B2**

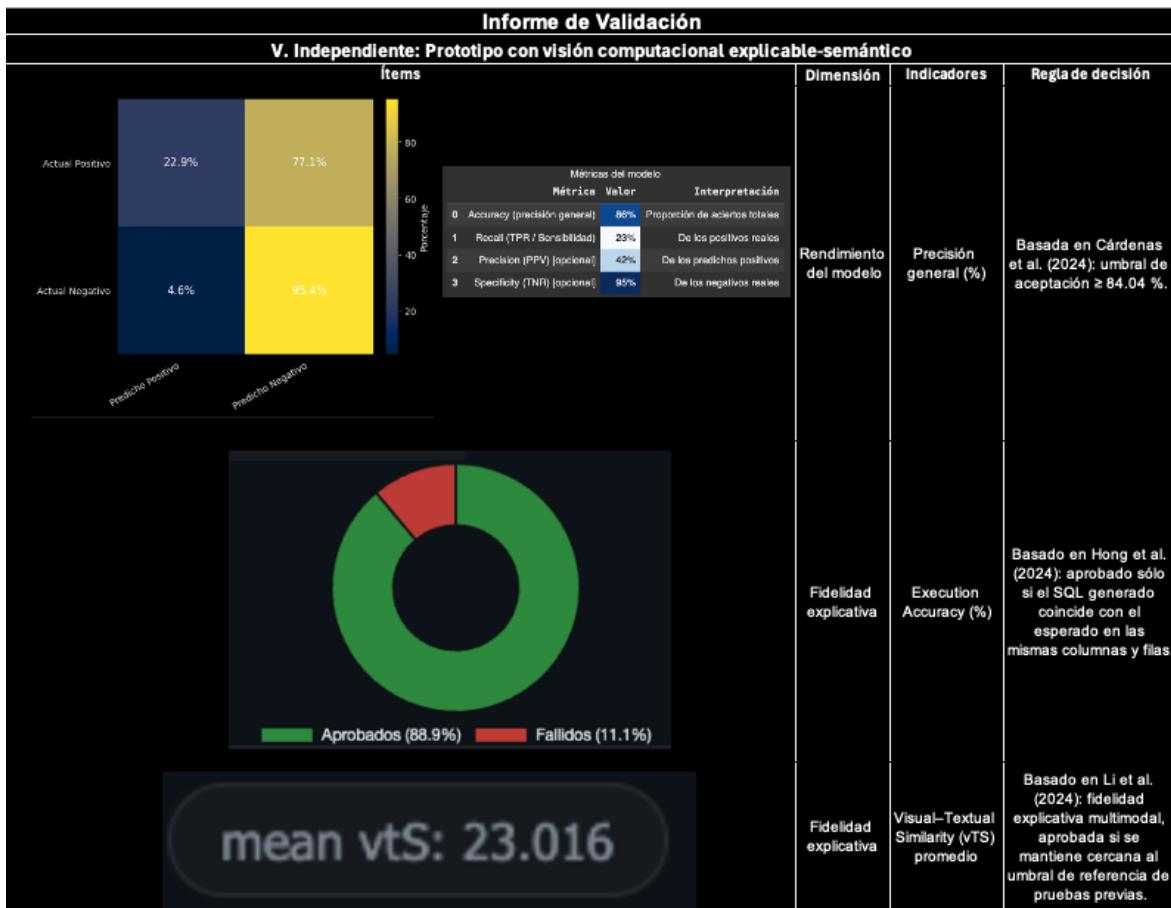
*Ficha Observación y sus ítems*

Variable	Ítem del formulario	Dimensión	Indicador
1. Nombre del archivo  Escriba su respuesta	1. <i>Nombre del archivo</i>	Identificación de la unidad de análisis	Código único de la imagen satelital
2. Corregimiento  <input type="radio"/> Amelia Denis de Icaza <input type="radio"/> Arnulfo Arias <input type="radio"/> Belisario Frías <input type="radio"/> Belisario Porras <input type="radio"/> José Domingo Espinar <input type="radio"/> Mateo Iturralde <input type="radio"/> Omar Torrijos <input type="radio"/> Rufina Alfaro <input type="radio"/> Victoriano Lorenzo	2. <i>Corregimiento</i>	Distribución espacial	Porcentaje % de acumulaciones por corregimiento
3. Cantidad de acumulaciones  Escriba su respuesta	3. <i>Cantidad de acumulaciones</i>	Cantidad de residuos	Número de acumulaciones Total de acumulaciones por Promedio de acumulaciones

*Nota.* Elaboración propia derivada de Cañete (2024).

**Figura B3**

*Informe de Validación*



*Nota.* Elaboración propia.

**Figura B4**

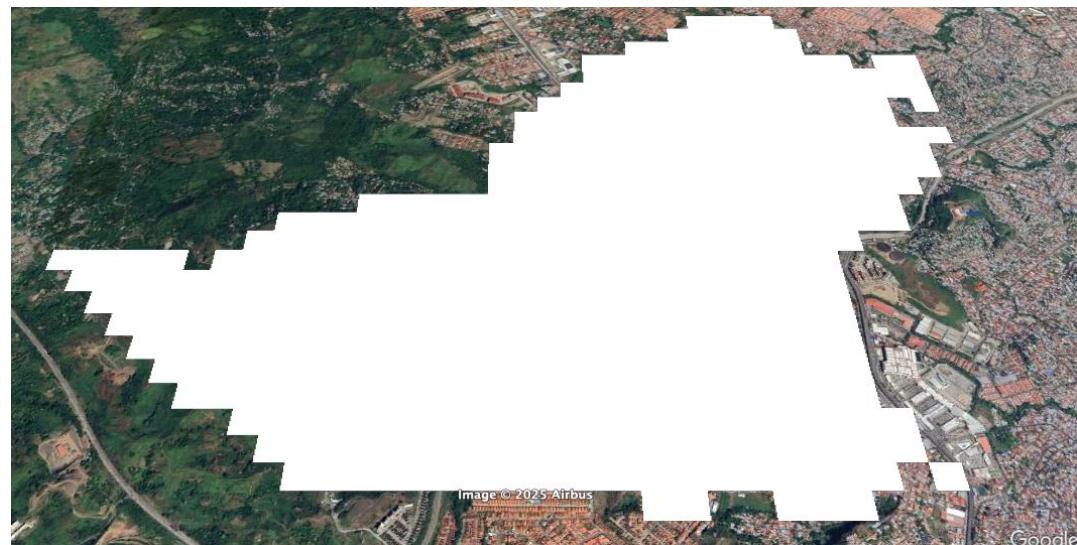
*Consultas Overpass para el Corregimiento Omar Torrijos*



*Nota.* Elaboración propia a partir de una consulta en Overpass Turbo (2025).

**Figura B5**

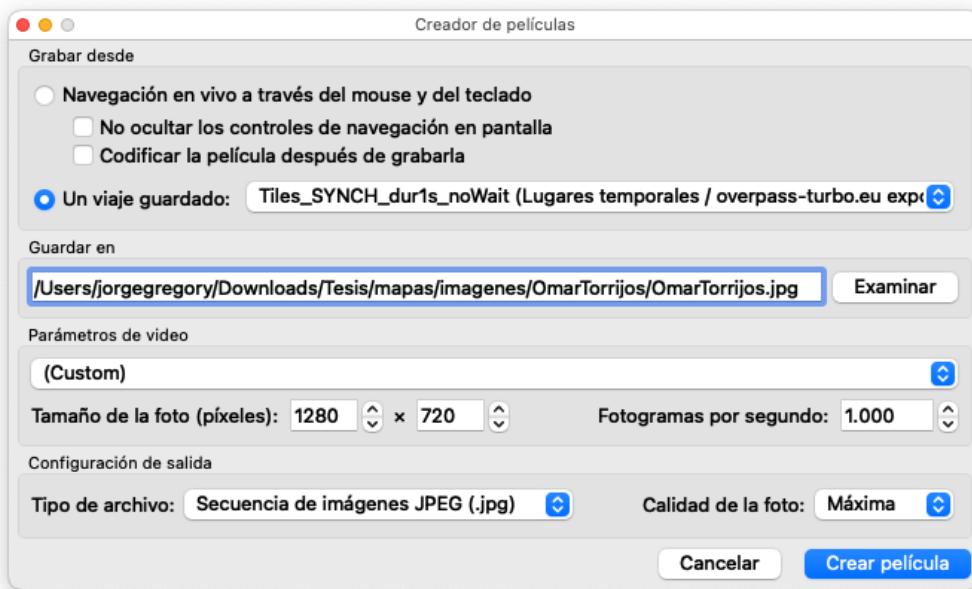
*Representación Gráfica del Contenido del Archivo de Omar Torrijos*



*Nota.* Captura de pantalla de un archivo KML visualizado en Google Earth Pro (2025).

## Figura B6

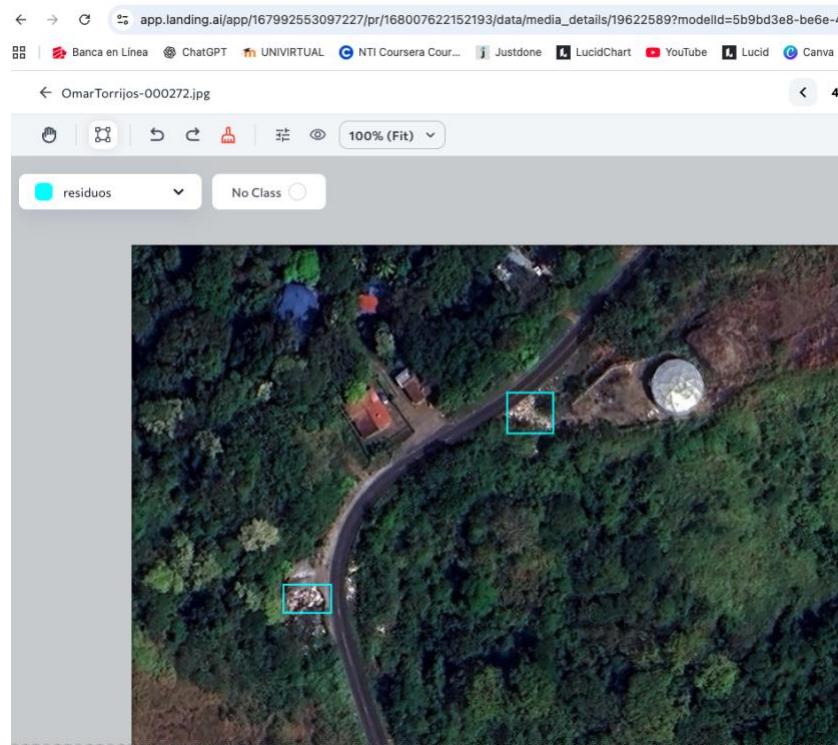
### Parámetros de Generación de Imágenes



Nota. Captura de pantalla del creador de películas en Google Earth Pro (2025).

## Figura B7

### Ejemplo de Etiquetado



Nota. Captura de pantalla del etiquetador de imágenes en LandingLens (2025).

## Figura B8

### Parámetros Augmentación de Imágenes con Python

```
# Pipeline de augmentación
transform = A.Compose([
    A.Rotate(limit=10, p=0.5),
    A.HorizontalFlip(p=0.5),
    A.RandomBrightnessContrast(p=0.8),
    A.HueSaturationValue(p=0.8),
    A.GaussianBlur(p=0.3),
    A.MotionBlur(p=0.2),
    A.ShiftScaleRotate(shift_limit=0.05, scale_limit=0.2, rotate_limit=0, p=0.5),
])

```

Copiar Editar

Nota. Captura de Pantalla de Jupyter Notebook local (2025).

## Figura B9

### Prompt en Memex

Diseña una interfaz web moderna, amigable y responsive que permita:

- Subir imágenes con drag & drop y mostrar el progreso.
- Visualizar un historial en tabla con opción de abrir un modal de detalle.
- Incluir un visor de imágenes con un canvas overlay: bounding box rojo grueso, borde blanco fino, etiqueta roja con score y un punto central amarillo.
- Agregar un chat estilo copiloto con:
  - Diseño de burbujas modernas.
  - Tablas con encabezados en español natural.
  - Scroll automático al último mensaje.
  - Mostrar imágenes anotadas solo cuando el usuario pida describir una imagen concreta.
- Estilo visual: modo oscuro estilo GitHub (fondos #0d1117, bordes #30363d, acentos #58a6ff).

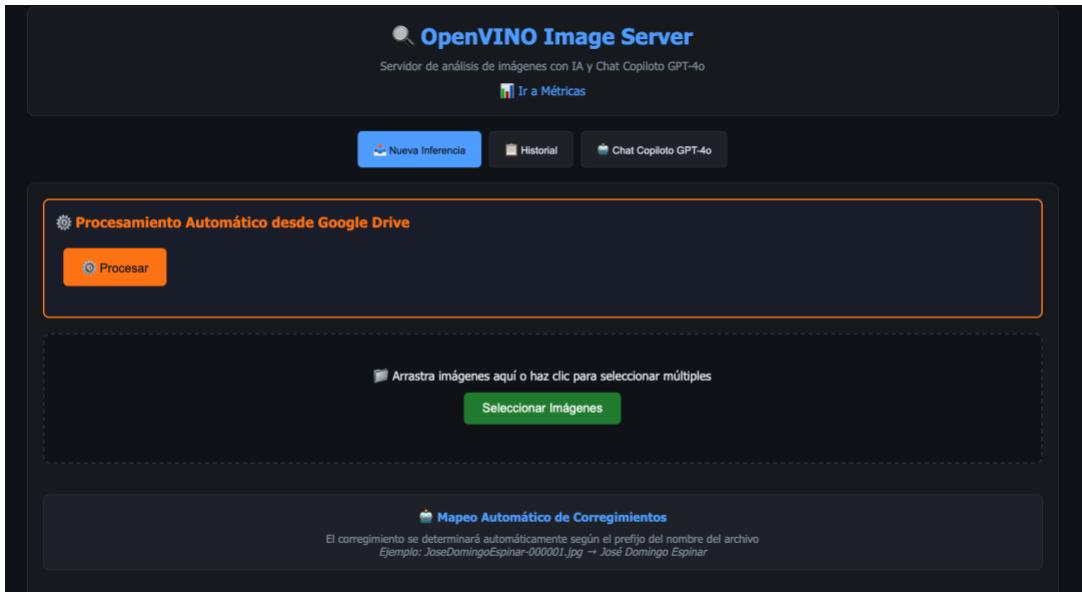
Necesito que armes el módulo "Chat Copiloto GPT-4o" dentro de nuestra SPA, cuidando tanto la experiencia como la integración técnica. Mantén un estado de conversación con el ID actual, la calidad seleccionada (arranca en full, permite mini), la cola de mensajes y si hay una petición en curso. Cada vez que el usuario envíe texto, valida que no venga vacío, crea un AbortController y llama a POST /chat con cabeceras Content-Type: application/json y Accept: text/event-stream, application/json, enviando este payload:

```
{
  "model": "gpt-4o-mini",
  "messages": [
    {
      "role": "system",
      "content": "Eres un **copiloto determinista** de una app de visión orientada a residuos...\n\n**Formato obligatorio:**\n"
    },
    {
      "role": "user",
      "content": "¿Puedes resumirme las últimas imágenes procesadas con detecciones?"
    }
  ],
  "response_format": {
    "type": "json_schema",
    "json_schema": {
      "name": "too_response",
      "schema": { ... esquema JSON usado por la app ... }
    }
  },
  "temperature": 0.1,
  "top_p": 0.95,
  "seed": 7
}
```

Nota. Captura de pantalla en Memex (2025).

## Figura B10

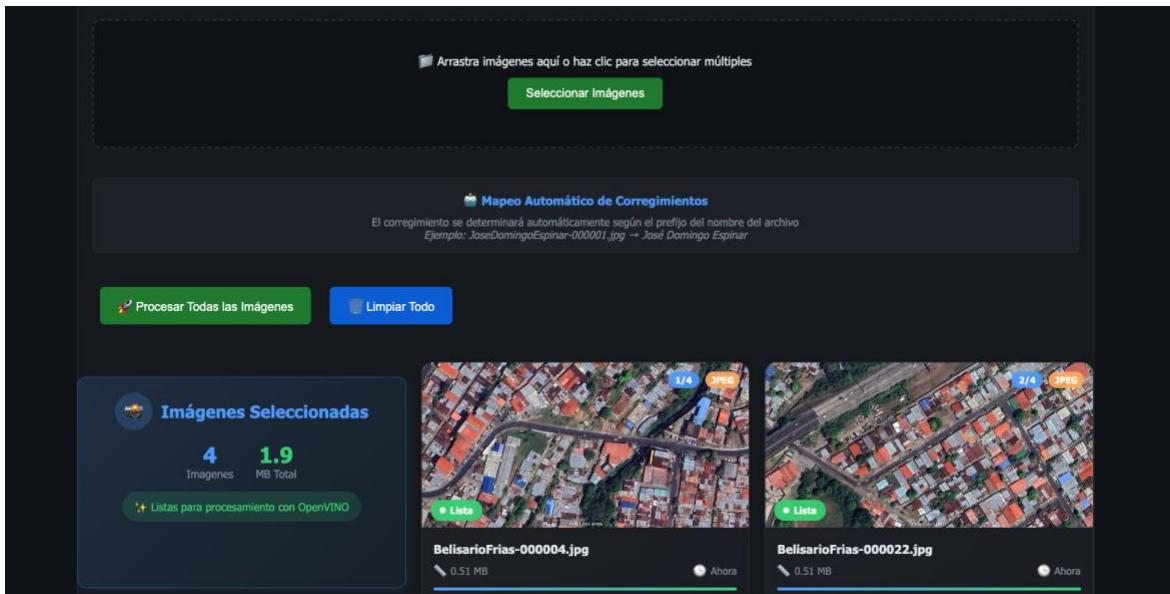
Panel de Nuevas Inferencias



Nota. Captura de pantalla del prototipo (2025).

## Figura B11

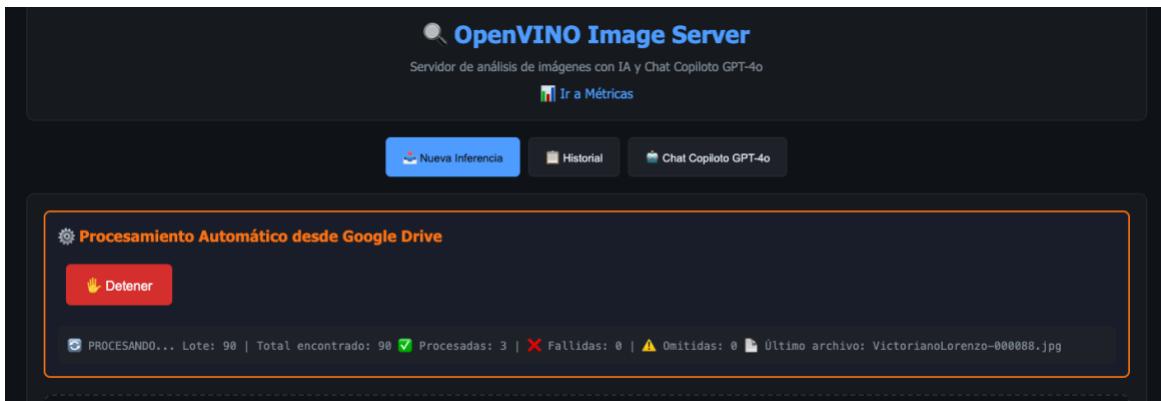
Carga de Imágenes desde la Computadora



Nota. Captura de pantalla del prototipo (2025).

**Figura B12**

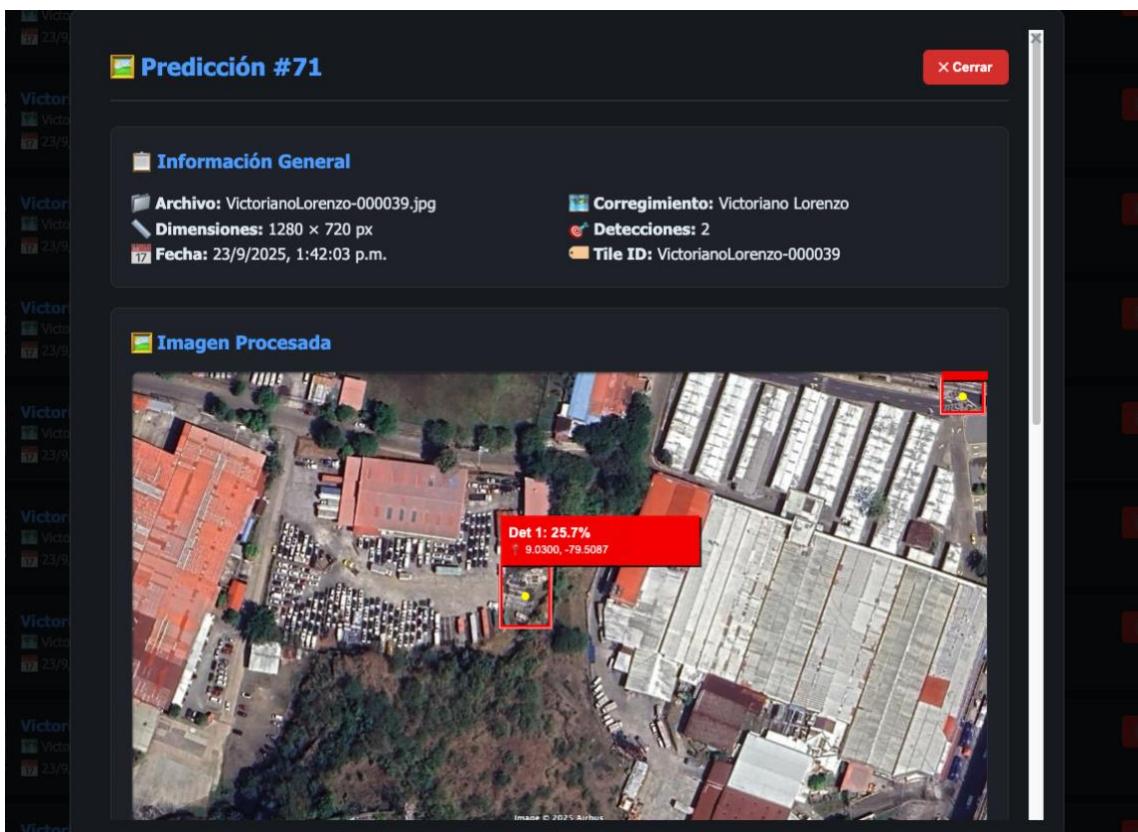
*Descarga de Imágenes desde Google Drive*



*Nota.* Captura de pantalla del prototipo (2025).

**Figura B13**

*Vista Previa de la Predicción*



*Nota.* Captura de pantalla del prototipo (2025).

**Figura B14**

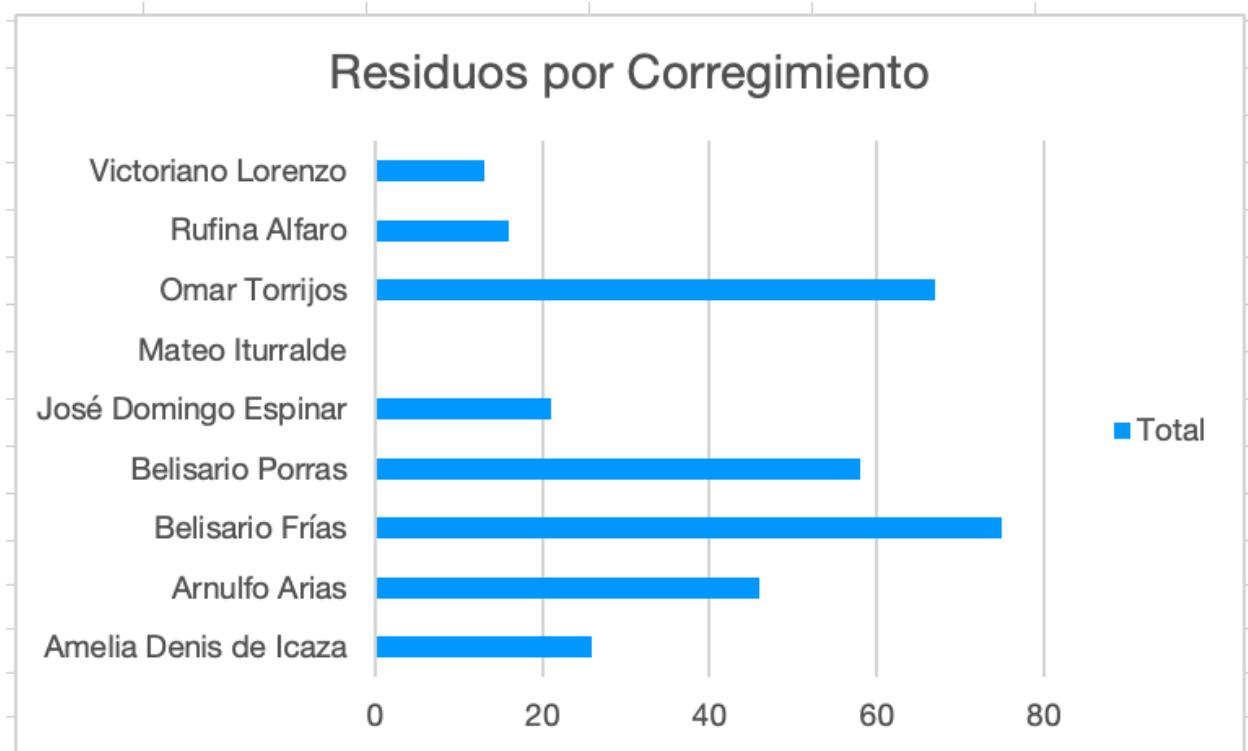
*Chat Copiloto GPT-4o Generando Respuesta Textual y Visual*



Nota. Captura de pantalla del prototipo (2025).

**Figura B15**

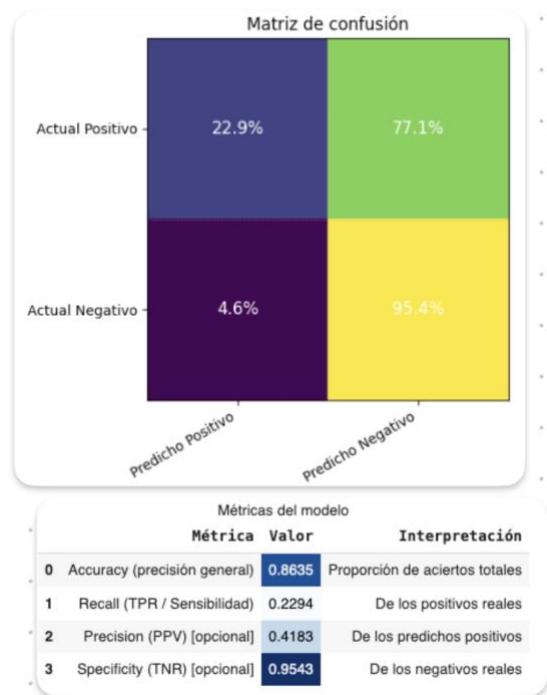
*Gráfico de Distribución de Residuos*



Nota. Elaboración propia en Excel (2025).

**Figura B16**

*Matriz de Confusión y Métricas relativas*



*Nota.* Elaboración propia a partir de Jupyter Notebook Local (2025).

**Figura B17**

*Prompts de Creación de Métricas de Fidelidad.*

**Prompt para crear la métrica Execution Accuracy**  
 Necesito implementar desde cero la métrica *Execution Accuracy* para nuestro módulo NLQ. La entrada serán pares pregunta → SQL\_gold guardados en un benchmark. La métrica debe ejecutar la consulta generada por el LLM, compararla con el SQL de referencia y entregar:

- porcentaje global de aciertos,
- número de preguntas correctas/incorrectas y fallas de ejecución,
- detalles por ítem (SQL generado, diff con el gold, errores),
- endpoints REST y un panel web con la tabla de resultados y export CSV.

Pliega la estructura de datos, la lógica de comparación y la UI para que luego pueda integrarla en la tesis."

**Prompt para crear la métrica vtS (Visual-Textual Similarity)**  
 "Quiero desarrollar la métrica *Visual-Textual Similarity* (*vtS*) usando CLIP para evaluar explicaciones sobre imágenes. Necesitamos:

1. Preparar un benchmark en CSV que recorra `image_preds` y agregue, por imagen, un texto 'gold' (con conteos de detecciones) y el bounding box.
2. Crear un pipeline que le pida a un LLM la explicación de cada imagen (incluye trampas aleatorias para simular errores).
3. Calcular el *vtS* entre la imagen y el texto (gold o LLM) y devolver mean, median, percentiles, N y cuántos superan el umbral.
4. Exponer todo vía API y un dashboard con gráficos, tabla, y exportables.  
 Diseña los modelos, la API y la UI para dejar la métrica lista y documentable en la tesis."

*Nota.* Captura de pantalla de Memex (2025).

### **Figura B18**

*Fórmula de Precisión para Modelos de Inteligencia Artificial*

$$\text{Precisión} = \frac{TP}{TP + FP}$$

**TP:** verdaderos positivos

**FP:** falsos positivos

Nota. Elaboración propia basado en LandingLens (2025).

### **Figura B19**

*Fórmula de Recall para Modelos de Inteligencia Artificial*

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

**TP:** verdaderos positivos

**FN:** falsos negativos

Nota. Elaboración propia basado en LandingLens (2025).

### **Figura B20**

*Formula de Mean Average Precision (mAP)*

$$\text{mAP} = \frac{1}{C} \sum_{c=1}^C \text{AP}_c$$

**mAP:** mean Average Precision

**C :** número total de clases

$\frac{1}{C}$  : factor de promedio

$\sum_{c=1}^C$  : suma sobre todas las clases.

**c :** índice de la clase

**AP<sub>c</sub>** : Average Precision de la clase c

Nota. Elaboración propia basado en LandingLens (2025).

## Figura B21

*Fórmula de Matriz de Confusión*

$$\begin{aligned} \text{TP} &= \sum_{i=1}^N \mathbf{1}[\hat{y}_i = 1 \wedge y_i = 1], \\ \text{FP: verdaderos positivos} &\quad \text{FP} = \sum_{i=1}^N \mathbf{1}[\hat{y}_i = 1 \wedge y_i = 0], \\ \text{FP: falsos positivos} &\quad \text{FN} = \sum_{i=1}^N \mathbf{1}[\hat{y}_i = 0 \wedge y_i = 1], \\ \text{FN: falsos negativos} &\quad \text{TN} = \sum_{i=1}^N \mathbf{1}[\hat{y}_i = 0 \wedge y_i = 0]. \\ \text{TN: verdaderos negativos} & \end{aligned}$$

Matriz (filas = **Real**, columnas = **Predicha**):

		$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	TN	FP	
$y = 1$	FN	TP	

*Nota.* Elaboración propia basado en LandingLens (2025).

## Figura B22

*Fórmula de Precisión General*

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

TP: verdaderos positivos

TN: verdaderos negativos

FP: falsos positivos

FN: falsos negativos

*Nota.* Elaboración propia en LandingLens (2025).

## Figura B23

*Fórmula de Execution Accuracy (EX)*

$$EX = \frac{1}{N} \sum_{n=1}^N \mathbf{1} \left[ \hat{V}_n = V_n \right]$$

*Nota.* Elaboración propia basada en Hong et al. (2024).

**Figura B24**

*Fórmula de Visual-Textual Similarity (VTS)*

$$TS = \frac{1 + \cos(\text{Emb}(T_{\text{pred}}), \text{Emb}(T_{\text{gt}}))}{2}$$

$$\text{vtS} = \frac{2 \text{ TS AP}}{\text{TS} + \text{AP}}$$

*Nota.* Elaboración propia basada en Li et al. (2024).