

Universidad del Valle de Guatemala

Facultad de ingeniería

Data Science

Catedrático: Luís Furlán



Proyecto 2.

Análisis Exploratorio

Nelson Eduardo García Bravatti 22434

Joaquín André Puente Grajeda 22296

José Antonio Mérida Castejón 201105

Guatemala, septiembre de 2025

Planteamiento inicial del problema:

a) Situación problemática

Las enfermedades transmitidas por mosquitos (p. ej., dengue, zika y chikunguña) requieren sistemas de vigilancia oportunos. Los métodos tradicionales son costosos y lentos; por ello, proyectos de ciencia ciudadana como Mosquito Alert permiten que la población reporte observaciones y fotos de mosquitos, que posteriormente son validadas por especialistas y utilizadas para la gestión de salud pública. Automatizar la validación de imágenes reduciría la carga de trabajo y aceleraría las decisiones de control vectorial.

b) Problema científico

¿Es posible entrenar un modelo de visión por computadora que detecte con precisión mosquitos (cumpliendo un umbral mínimo de IoU de 0.75) y clasifique de forma fiable su género/especie en imágenes reales capturadas por la ciudadanía, a pesar del fuerte desbalance de clases del dataset? La solución debe maximizar el Macro F1 en clasificación y mantener la calidad de detección exigida por el reto.

c) Objetivos

Objetivo general:

Desarrollar y evaluar un pipeline de aprendizaje profundo para detección y clasificación de mosquitos en imágenes del proyecto Mosquito Alert, cumpliendo el umbral de $\text{IoU} \geq 0.75$ y optimizando el Macro F1 en el conjunto de evaluación.

Objetivos específicos (medibles y alcanzables):

- Caracterizar el dataset (10,700 imágenes reales con cajas y seis clases: *Aedes aegypti*, *Aedes albopictus*, *Anopheles*, *Culex*, *Culiseta* y *Aedes japonicus/koreicus*), cuantificando el desbalance y el tamaño relativo del objeto.
- Diseñar el preprocesamiento y estrategias contra el desbalance y el tamaño pequeño del objeto (p. ej., recortes guiados por bbox, aumento de datos, muestreo estratificado/ponderado).
- Entrenar y comparar al menos dos arquitecturas de detección/clasificación de objetos (p. ej., un detector de una etapa y otro de dos etapas) y seleccionar el mejor modelo según Macro F1 (clasificación) y cumplimiento de $\text{IoU} \geq 0.75$ (detección).

- Documentar y reproducir la solución (código, pesos y guía de inferencia) para evaluación ciega en el conjunto de prueba, siguiendo el formato del challenge.

Investigación preliminar:

El dataset del Mosquito Alert Challenge contiene imágenes anotadas de seis clases: *Aedes aegypti*, *Aedes albopictus*, *Anopheles* (género), *Culex* (género), *Culiseta* (género) y el complejo *Aedes japonicus/koreicus*. Estas clases incluyen los principales vectores de arbovirosis y malaria en humanos.

A nivel global, las enfermedades transmitidas por vectores representan >17% de las enfermedades infecciosas y causan >700 000 muertes anuales; el dengue batió récords recientes en América con >13 millones de casos en 2024.

Especies y géneros del dataset:

Aedes aegypti (especie)

Rasgos: patrón en “lira” blanco en el tórax; cría en recipientes artificiales; pica sobre todo de día.
Vectores: dengue, chikunguña, Zika y fiebre amarilla.

Aedes albopictus (especie)

Rasgos: franja blanca única en el centro del tórax (“tigre asiático”); tolera climas más fríos que *Ae. aegypti*. Vectores: chikunguña y dengue (entre otros); menos antropofílico que *Ae. aegypti*.

Aedes japonicus/Aedes koreicus (complejo)

Rasgos: mosquitos invasores de zonas templadas; morfológicamente similares, por eso se agrupan como complejo. Vectores (evidencia): competencia vectorial demostrada en laboratorio para CHIKV/ZIKV/WNV, pero la relevancia epidemiológica aún es incierta en campo.

Anopheles (género)

Rasgos: palpos tan largos como el probóscide y postura de reposo inclinada; activos sobre todo anochecer/amanecer. Vectores: únicos transmisores de malaria en humanos (hembras).

Culex (género)

Rasgos: pican principalmente de noche; muchas especies prefieren aves. Vectores: virus del Nilo Occidental (WNV) y encefalitis (SLE/EEE/WEE) según la especie y región; Cx. pipiens, Cx. quinquefasciatus y Cx. tarsalis son claves en WNV.

Culiseta (género)

Rasgos: varias especies son de climas fríos; algunas de mayor tamaño. Vectores: Culiseta melanura sostiene el ciclo enzoótico de encefalitis equina del Este (EEE).

Impacto en salud pública y sociedad:

Carga sanitaria: además del dato global anterior, la región de las Américas vivió la mayor epidemia de dengue registrada (2024), con interrupción de servicios de salud y costos sociales elevados.

Costos económicos: las especies invasoras Aedes (especialmente aegypti y albopictus) acumulan costos reportados de al menos US\$ 94.7 mil millones en 45 años (manejo, pérdidas y daños).

Rol ambiental y consideraciones ecológicas:

Funciones ecológicas: los mosquitos (machos y muchas hembras fuera de la oviposición) se alimentan de néctar y pueden actuar como polinizadores (p. ej., Aedes communis en orquídeas boreales); además, sus larvas y adultos forman parte de cadenas tróficas acuáticas y terrestres.

Control y resistencia: el uso extensivo de insecticidas ha impulsado resistencia en Aedes y Anopheles; la OMS recomienda redes tratadas con nuevos ingredientes (p. ej., clorfenapir) en áreas con resistencia a piretroides.

Análisis inicial del problema y los datos disponibles: