

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:

Análisis inicial del problema y los datos disponibles.