

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

**Detección Automática de Escarabajos Mariquita en Imágenes de
la Vida Silvestre**

Israel Sebastián Granizo Freire

Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniería en Ciencias de la Computación

Quito, 09 de mayo de 2025

Universidad San Francisco de Quito USFQ
Colegio de Ciencias e Ingeniería

**HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA**

**Detección Automática de Escarabajos Mariquita en Imágenes de la
Vida Silvestre**

Israel Sebastián Granizo Freire

Nombre del profesor, Título académico

Noel Pérez, Ph.D.

Quito, 09 de mayo de 2025

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Israel Sebastián Granizo Freire

Código: 00212771

Cédula de identidad: 1804358198

Lugar y fecha: Quito, 09 de mayo de 2025

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>.

RESUMEN

Este estudio explora la aplicación de técnicas de aprendizaje profundo para la detección automatizada de insectos en entornos agrícolas mediante visión por computadora. El control de insectos es esencial para la protección de cultivos, ya que una identificación oportuna y precisa de especies dañinas puede reducir significativamente las pérdidas agrícolas y evitar el uso excesivo de pesticidas. La detección temprana permite a los agricultores aplicar medidas de control específicas, mejorando la calidad del rendimiento y promoviendo la sostenibilidad.

En esta investigación se evaluaron múltiples versiones del modelo de detección de objetos YOLO (You Only Look Once), entrenadas con un conjunto de datos personalizado que incluye imágenes etiquetadas de insectos, con un enfoque en diferenciar especies dañinas de las benignas. El conjunto de datos fue proporcionado por iNaturalist Ciencia Ciudadana y contiene 2899 imágenes de mariquitas (Coccinellidae) capturadas en diversos entornos agrícolas de Ecuador, Colombia, Chile, Perú y Bolivia. Las imágenes, tomadas en condiciones naturales, presentan variaciones de iluminación, fondos y orientaciones, lo que permitió entrenar modelos robustos y generalizables a escenarios reales.

Se ajustaron y probaron versiones pequeñas del modelo YOLO bajo los mismos parámetros, y se utilizaron pruebas estadísticas como Wilcoxon y Mann-Whitney U para comparar su desempeño. El modelo YOLOv11 demostró ser el más eficaz, identificando correctamente los objetos en un promedio de 15 milisegundos. Logró una precisión promedio (mAP@50) de 86.8%, mAP@50–95 de 55.01%, precisión de 97.5% y recall de 75.8%. Estos resultados reflejan la efectividad del modelo propuesto para

detectar mariquitas en imágenes naturales, con potencial para aplicaciones reales en monitoreo agrícola.

Palabras clave: Detección de Insectos, Visión por Computadora, YOLO, Escarabajos Mariquitas, Aprendizaje Profundo.

ABSTRACT

This paper aims to explore the application of deep learning techniques for the automated detection of insects in agricultural environments using computer vision. Insect control is a crucial aspect of crop protection, as timely and accurate identification of harmful species can significantly reduce crop damage and minimize the overuse of pesticides. Early detection helps farmers implement targeted control measures, improving yield quality and sustainability. In this study, we evaluated multiple versions of the YOLO (You Only Look Once) object detection model, each trained on a custom dataset of labeled insect images, with a specific focus on distinguishing between harmful and benign species. For this purpose, we used a dataset provided by iNaturalist Ciencia Ciudadana, which contains 2899 images of ladybird beetles (Coccinellidae) collected from various agricultural environments across Ecuador, Colombia, Chile, Peru, and Bolivia. The dataset includes images captured in natural settings, showcasing the beetles in the wild with varying lighting, background conditions, and orientations. This diversity in the data allowed us to train and refine the core object detection methods, ensuring that the models can generalize effectively to real-world conditions. Through this work, we aim to contribute to the development of more efficient, accurate, and cost-effective solutions for insect monitoring in agriculture. We fine-tuned and tested different YOLO models with the small architecture with the same set of parameters. All models were compared and analyzed using statistical testings such as the Wilcoxon test and the Mann-Whitney U test. We observed that the YOLOv11 model correctly identified the desired objects at an average speed of 15 ms. The proposed method achieved an average precision value of (mAP@50) of 86.8%, (mAP@50-95) 55.01%, precision of 97.5% and a recall of 75.8%. These results depict how the proposed model worked effectively and was able to efficiently identify and detect ladybird beetles in

wildlife images, which can be employed for realistic applications while trying to identify and detect ladybird beetles in wildlife images.

Keywords: Wildlife Insects, Computer Vision, Deep Learning, YOLO Detector, Ladybird Beetles Detection.

TABLA DE CONTENIDO

1. Introducción.....	12
2. Desarrollo del Tema.....	15
2.1. Métodos y materiales.....	15
2.1.1. Base de datos.....	15
2.1.2. Método propuesto.....	15
2.2. Setup Experimental.....	19
2.2.1 Preparación del Dataset.....	19
2.2.2 Conjuntos de entrenamiento, validación y prueba.....	19
2.2.3. Optimización de modelos.....	20
2.2.4. Métricas de evaluación y regla de selección.....	20
3. Resultados y discusión.....	21
3.1. Rendimiento en la fase de entrenamiento.....	22
3.2. Evaluación del rendimiento en la fase de prueba.....	24
3.3. Comparación con el estado del arte.....	27
4. Conclusiones y Trabajo Futuro.....	29
Referencias bibliográficas.....	30

ÍNDICE DE TABLAS

Tabla #1. Resultados de detección para cada versión pequeña de YOLO utilizando un método de validación cruzada de diez particiones.....	22
---	----

ÍNDICE DE FIGURAS

Figura #1. Flujo de trabajo experimental del método propuesto.....	17
Figura #2. Pérdida de entrenamiento y validación.....	24
Figura #3. Comparación visual entre imágenes de mariquitas pequeñas, medianas y diminutas.....	26
Figura #4. Escarabajos mariquitas que presentaron desafíos en la detección.....	27

1. Introducción

Los insectos representan el grupo de animales más diverso del planeta y desempeñan funciones esenciales en los ecosistemas, tales como la polinización, el reciclaje de nutrientes y el control biológico de plagas. A pesar de su importancia ecológica, las poblaciones de insectos están experimentando un marcado declive a nivel global, lo que hace que su monitoreo e identificación precisa sean más cruciales que nunca. Sin embargo, la identificación tradicional de insectos continúa siendo una tarea compleja debido al gran número de especies, la variabilidad morfológica y la escasez de taxónomos especializados.

Entre los insectos, los escarabajos mariquitas (familia Coccinellidae) tienen una relevancia tanto ecológica como agrícola. Son ampliamente reconocidos por su coloración llamativa y por su papel como depredadores naturales de plagas agrícolas comunes (Dixon, 2000). Aunque se han descrito miles de especies, su diversidad sigue siendo poco comprendida en muchas regiones, especialmente en el Neotrópico. A ello se suma que algunas especies han sido introducidas en diversas partes del mundo como agentes de control biológico, con el riesgo de convertirse en invasoras y alterar los ecosistemas locales (Simberloff y Stiling, 1996).

Detectar tanto especies nativas como invasoras de mariquitas en entornos naturales es fundamental para la conservación de la biodiversidad y el monitoreo agrícola. No obstante, la identificación manual requiere mucho tiempo, es propensa a errores y no resulta escalable para grandes volúmenes de datos. En este contexto, los enfoques emergentes basados en aprendizaje automático y la participación comunitaria mediante ciencia ciudadana abren nuevas posibilidades para desarrollar herramientas que mejoren significativamente el descubrimiento e identificación de especies de insectos. Además,

estas herramientas pueden favorecer el reconocimiento temprano de especies introducidas, apoyando estrategias de monitoreo, gestión y control de invasiones biológicas (Venegas et al., 2021).

Los métodos de detección basados en aprendizaje automático, tanto superficial como profundo, ofrecen un valioso apoyo a los expertos encargados de detectar diferentes tipos de insectos con fines biológicos o agrícolas (Xiong et al., 2024). En particular, las técnicas de aprendizaje profundo como YOLO (Ahmad et al., 2022; Kittichai et al., 2021; Mamdouh y Khattab, 2021; N. Wang et al., 2025), redes neuronales convolucionales (CNN) (Teixeira et al., 2023; X. Wang et al., 2023), máquinas de vectores de soporte (SVM) (Gondal y Khan, 2015), transformers visuales (ViT) (Fu et al., 2024), métodos de saliencia (Nanni et al., 2020) y redes neuronales artificiales (ANN) (Teixeira et al., 2023), han sido empleadas en distintos experimentos para abordar la tarea de detección de insectos. Si bien los avances tecnológicos han mejorado el reconocimiento automatizado de especies, sigue existiendo una atención limitada hacia la detección de mariquitas en entornos naturales (Bjerge et al., 2023; Khujamatov et al., 2025; Walter et al., 2021). Gran parte de la investigación existente no ha abordado de forma exhaustiva los retos que implica localizar estos insectos en escenarios complejos del mundo real, donde factores como la vegetación de fondo, la iluminación variable y la diversidad de hábitats dificultan la detección. Los métodos tradicionales, como la inspección manual por expertos, siguen utilizándose ampliamente, pero son intensivos en tiempo y susceptibles a errores humanos, especialmente al tratar grandes volúmenes de datos o realizar trabajo de campo prolongado.

Por lo tanto, proponemos explorar tres arquitecturas ligeras de modelos de detección de objetos YOLO —YOLOv10, YOLOv11 y YOLOv12— con el objetivo de maximizar la detección de mariquitas en ambientes silvestres. YOLO es especialmente adecuado para

aplicaciones en tiempo real debido a su capacidad para procesar imágenes completas en una sola pasada, identificando y localizando objetos de manera eficiente. Al aprovechar su rapidez y precisión, esta técnica permite detectar de forma directa ubicaciones potenciales de mariposas en imágenes del entorno natural, incluso bajo condiciones ambientales diversas. Esta estrategia reduce considerablemente el costo computacional frente a técnicas tradicionales como las ventanas deslizantes, manteniendo un alto rendimiento en la detección. En consecuencia, ofrece una solución práctica y escalable para apoyar el monitoreo ecológico y la investigación en biodiversidad en hábitats reales y dinámicos.

El resto del artículo se organiza de la siguiente manera: en la sección Materiales y Métodos se describen en detalle la base de datos utilizada, las arquitecturas de los modelos, y los procedimientos de entrenamiento y evaluación. La sección de Resultados y Discusión presenta los hallazgos experimentales, las métricas de rendimiento y el análisis del comportamiento del modelo. Finalmente, la sección de Conclusiones y Trabajo Futuro resume los principales resultados y plantea posibles líneas para ampliar las capacidades del sistema.

2. Desarrollo del Tema

2.1. Métodos y materiales.

2.1.1. Base de datos.

Utilizamos una base de datos de acceso público proporcionada por el sitio web de Ciencia Ciudadana iNaturalist (<https://www.inaturalist.org/>), que contiene 2899 imágenes de escarabajos mariquitas (Coccinellidae) recolectadas en Ecuador, Colombia, Chile, Perú y Bolivia. Estas imágenes capturan numerosas especies de mariquitas en una variedad de posturas, condiciones de iluminación y entornos naturales. Además, la base de datos empleada está clasificada como *research-grade*, lo que significa que todas las imágenes han sido verificadas y validadas por expertos taxonómicos a través de la plataforma iNaturalist. Esta validación por parte de especialistas garantiza un alto nivel de fiabilidad y precisión en el etiquetado de las especies, lo que hace que la base de datos sea particularmente adecuada para aplicaciones científicas y el monitoreo ecológico.

2.1.2. Método propuesto.

El método propuesto se basa en la implementación de tres arquitecturas ligeras del modelo YOLO (You Only Look Once): YOLOv10, YOLOv11 y YOLOv12, con el objetivo de maximizar la detección de escarabajos mariquitas en entornos naturales. Optamos por el enfoque YOLO frente a alternativas como redes neuronales artificiales (ANN), redes neuronales convolucionales (CNN) sin capacidades de localización y transformadores de visión (ViT), debido al equilibrio único que ofrece entre precisión, velocidad de inferencia y eficiencia de implementación. A diferencia de los detectores

de objetos multietapa o de los modelos basados únicamente en clasificación —que requieren procesamiento adicional para la localización—, YOLO es un detector single-shot que realiza detección y clasificación de objetos en una sola pasada, lo que lo hace ideal para aplicaciones en tiempo real en entornos dinámicos. Su capacidad para detectar múltiples objetos de diferentes escalas dentro de una misma imagen reduce significativamente la necesidad de mecanismos complejos de preprocesamiento o de ventanas deslizantes, permitiendo resultados más rápidos y consistentes en condiciones de campo. La representación gráfica del flujo de trabajo propuesto con el modelo base YOLOv10 se muestra en la Fig. 1.

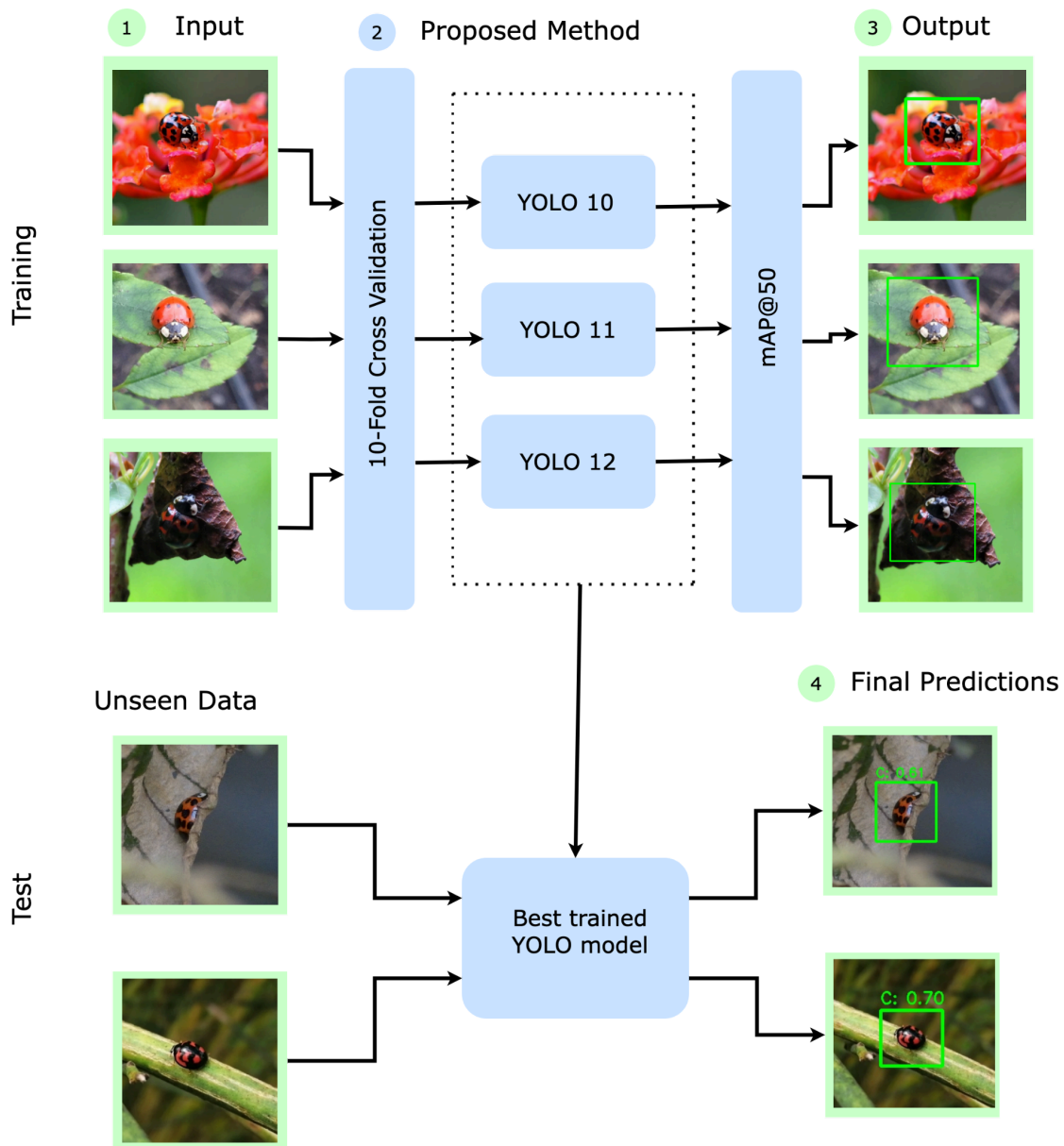


Figura 1. Flujo de trabajo experimental del método propuesto.

A continuación, se presenta una breve descripción de los modelos utilizados:

- **YOLOv10:** Es nuestro modelo base y se caracteriza por una arquitectura simplificada y compacta que prioriza la eficiencia computacional sin comprometer la capacidad de detección. Su estructura sigue un esquema codificador-decodificador típico, comenzando con una *backbone* de capas convolucionales apiladas, seguida de una etapa de *spatial pyramid pooling* y una

cabeza de detección. La *backbone* integra conexiones *Cross Stage Partial (CSP)*, que dividen los mapas de características y los fusionan posteriormente, mejorando el flujo del gradiente y reduciendo la redundancia. Cada bloque convolucional está seguido por normalización por lotes y funciones de activación, comúnmente *SiLU* o *ELU*, para facilitar una convergencia más rápida. Además, incorpora una red de agregación de caminos (*Path Aggregation Network, PAN*) en el *neck*, lo que permite una fusión multiescala de características, mejorando la detección de mariposas pequeñas en contextos variados de imagen (A. Wang et al., 2024).

- **YOLOv11:** Es una versión mejorada de YOLOv10, que introduce componentes arquitectónicos refinados para potenciar la detección de objetos pequeños y parcialmente ocultos. Las principales mejoras incluyen una fusión de características basada en BiFPN, mecanismos de atención como CBAM y *coordinate attention*, y generación dinámica de *anchor boxes* mediante *k-means clustering*. Estas mejoras incrementan la precisión espacial y la exactitud de los cuadros delimitadores, manteniendo al mismo tiempo el rendimiento en tiempo real y un diseño ligero, adecuado para despliegues en campo (Hidayatullah et al., 2025; Khanam y Hussain, 2024).
- **YOLOv12:** Introduce una arquitectura centrada en mecanismos de atención que mejora el rendimiento de detección sin sacrificar la capacidad de inferencia en tiempo real. Su principal innovación es el uso de *area attention*, que segmenta los mapas de características en regiones iguales, permitiendo al modelo enfocarse de manera eficiente en las zonas relevantes. Este enfoque reduce la complejidad computacional de cuadrática a lineal, lo que permite una extracción de características más rápida y precisa. Además, incorpora redes de agregación

de capas eficientes con conexiones residuales (Residual Efficient Layer Aggregation Networks, R-ELAN), que emplean conexiones residuales a nivel de bloque y técnicas de escalado para mejorar el flujo del gradiente y la agregación de características. La integración de FlashAttention optimiza aún más los patrones de acceso a memoria, reduciendo la latencia durante la inferencia (Alif y Hussain, 2025).

2.2. Setup Experimental

2.2.1 Preparación del Dataset

Utilizamos herramientas de anotación como Computer Vision Annotation Tool (CVAT) para dibujar cuadros delimitadores alrededor de cada escarabajo mariquita presente en las imágenes, manteniendo una distancia de 10 píxeles (determinada empíricamente) desde el contorno del insecto. A cada anotación se le asignó una etiqueta de clase de manera consistente. Posteriormente, todas las anotaciones fueron exportadas en un formato compatible con el entrenamiento de modelos YOLO.

2.2.2 Conjuntos de entrenamiento, validación y prueba.

Utilizamos una división tradicional de los datos en una proporción de 70 % para entrenamiento y 30 % para prueba, con el fin de garantizar una adecuada variación de datos durante el proceso de aprendizaje del método propuesto. De esta manera, el conjunto de entrenamiento alimentó un esquema de validación cruzada estratificada de diez pliegues para conformar subconjuntos de entrenamiento y validación. El conjunto de prueba externo se reservó para evaluar la capacidad de generalización de los modelos seleccionados sobre datos no vistos.

2.2.3. Optimización de modelos.

Para todos los modelos YOLO, se realizaron entrenamientos durante 200 épocas con un tamaño de lote de 16, empleando un mecanismo de detención anticipada de 20 épocas. Esto significa que el entrenamiento se detenía si no se observaba mejora en las métricas de validación después de 20 épocas consecutivas. Se utilizó el optimizador AdamW, el cual combina los beneficios del optimizador Adam con regularización por decaimiento de pesos (weight decay) para mejorar la capacidad de generalización.

Además, se aplicaron operaciones de aumento de datos en tiempo real (on-the-fly), como rotaciones aleatorias de hasta 45 grados, transformaciones de escala con un factor de 0.4 y volteo vertical con una probabilidad del 25 % (flipud: 0.25), con el objetivo de incrementar la variabilidad de los datos durante el proceso de entrenamiento.

Adicionalmente, se optimizó el umbral de confianza para las pruebas de validación en un rango de 0.5 a 0.9 con incrementos de 0.1 unidades.

2.2.4. Métricas de evaluación y regla de selección

Se calcularon las métricas $mAP@50$, $mAP@50-95$, precisión (PRE) y recall (REC) para evaluar el desempeño del método propuesto en la tarea de detección. Aunque la métrica $mAP@50-95$ ofrece una visión general más completa del rendimiento del modelo al considerar múltiples umbrales de IoU, se empleó principalmente la métrica $mAP@50$ como indicador principal para el análisis y discusión de los resultados obtenidos. Esta elección se fundamenta en las características particulares de nuestra tarea, cuyo objetivo es detectar de forma confiable la presencia y ubicación general de las mariquitas, más que lograr una localización extremadamente precisa.

Dado que el $\text{mAP}@50$ utiliza un único umbral de intersección sobre unión (IoU) de 0.5, proporciona una medida más permisiva e interpretable del éxito en la detección, lo cual lo hace especialmente adecuado para aplicaciones prácticas en monitoreo biológico, donde una detección aproximada suele ser suficiente (Ultralytics, 2023).

Además, se aplicó la prueba de rangos con signo de Wilcoxon con un valor de alfa de 0.05 para comparar estadísticamente el rendimiento de los modelos basados en $\text{mAP}@50$.

La selección del resultado del método propuesto se realiza según la siguiente regla: se elige el modelo que maximice la métrica $\text{mAP}@50$ y, en caso de empate en el rendimiento, se selecciona aquel que utilice el mayor umbral de confianza.

3. Resultados y discusión

Evaluamos un total de 15 modelos de detección derivados de tres versiones de YOLO con diferentes umbrales de confianza, utilizando una estrategia de validación cruzada de diez particiones (ten-fold cross-validation). Los resultados obtenidos se presentan en la Tabla 1.

Tabla #1. Resultados de detección para cada versión pequeña de YOLO utilizando un método de validación cruzada de diez particiones.

Model	Parameters	Threshold	mAP50(SD)	mAP50-95(SD)	Precision(SD)	Recall(SD)	U-Test
YOLOv10	p=20 b=16 e=200 o=AdamW	0.5	0.881(0.048)	0.573(0.042)	0.947(0.021)	0.793(0.088)	P < 0.05
		0.6	0.841(0.057)	0.554(0.045)	0.970(0.016)	0.704(0.109)	P < 0.05
		0.7	0.782(0.071)	0.525(0.050)	0.981(0.009)	0.580(0.139)	P < 0.05
		0.8	0.683(0.082)	0.478(0.051)	0.986(0.014)	0.381(0.158)	P < 0.05
		0.9	0.461(0.162)	0.353(0.127)	0.890(0.313)	0.032(0.028)	P < 0.05
YOLOv11		0.5	0.918(0.046)*	0.585(0.041)	0.956(0.021)	0.855(0.084)	—
		0.6	0.876(0.064)	0.567(0.046)	0.979(0.013)	0.762(0.124)	P = 0.053
		0.7	0.714(0.097)	0.482(0.060)	0.996(0.007)	0.432(0.194)	P < 0.05
		0.8	0.417(0.221)	0.292(0.162)	0.800(0.442)	0.034(0.046)	P < 0.05
		0.9	0.050(0.159)	0.040(0.127)	0.100(0.316)	0.000(0.0001)	P < 0.05
YOLOv12		0.5	0.913(0.041)	0.557(0.036)	0.949(0.032)	0.850(0.074)	P = 0.057
		0.6	0.874(0.054)	0.559(0.040)	0.975(0.015)	0.763(0.104)	P < 0.05
		0.7	0.720(0.083)	0.481(0.046)	0.990(0.011)	0.449(0.167)	P < 0.05
		0.8	0.432(0.229)	0.310(0.165)	0.797(0.420)	0.066(0.064)	P < 0.05
		0.9	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)	P < 0.05

SD: puntuación de desviación estándar; o: optimizador; p: paciencia; b: lote (batch); e: épocas; el valor subrayado indica el mayor puntaje de mAP50 (el pivote de comparación)

estadística con $\alpha = 0.05$); los valores en negrita corresponden al mejor modelo seleccionado; * representa el pivote estadístico.

3.1. Rendimiento en la fase de entrenamiento

A partir de la Tabla 1, se observa que la configuración con mejor desempeño fue YOLOv11 con un umbral de confianza de 0.5, alcanzando un $mAP@50$ de 0.918 (± 0.046), $mAP@50-95$ de 0.585 (± 0.041), precisión de 0.956 (± 0.021) y recall de 0.855 (± 0.084). Este modelo se utiliza como referencia para las comparaciones posteriores.

Según la prueba U de Mann-Whitney, YOLOv11 con umbral 0.6 y YOLOv12 con umbral 0.5 no presentaron diferencias estadísticamente significativas respecto al modelo pivote ($p > 0.05$), lo que indica un rendimiento similar.

El comportamiento de los modelos varió en función del umbral de confianza. Los umbrales de 0.5 y 0.6 ofrecieron mejores resultados, mientras que los umbrales más altos (0.8 y 0.9) redujeron considerablemente el recall, volviendo las detecciones más conservadoras. Este fenómeno es esperable en imágenes biológicas, donde los insectos pueden aparecer pequeños o parcialmente ocultos, y umbrales bajos permiten recuperar verdaderos positivos.

El proceso de entrenamiento de los modelos YOLO se desarrolló de manera estable, con una convergencia suave y sin señales de inestabilidad o degradación del rendimiento. Como se muestra en la Tabla 1, los modelos alcanzaron resultados sólidos en todos los umbrales evaluados, con valores consistentemente altos de $mAP@50$, precisión y recall, lo cual indica un desempeño confiable en la detección. Cabe destacar que las métricas se mantuvieron robustas incluso con umbrales más estrictos, lo que demuestra la

capacidad de los modelos para mantener la calidad de detección bajo condiciones más exigentes.

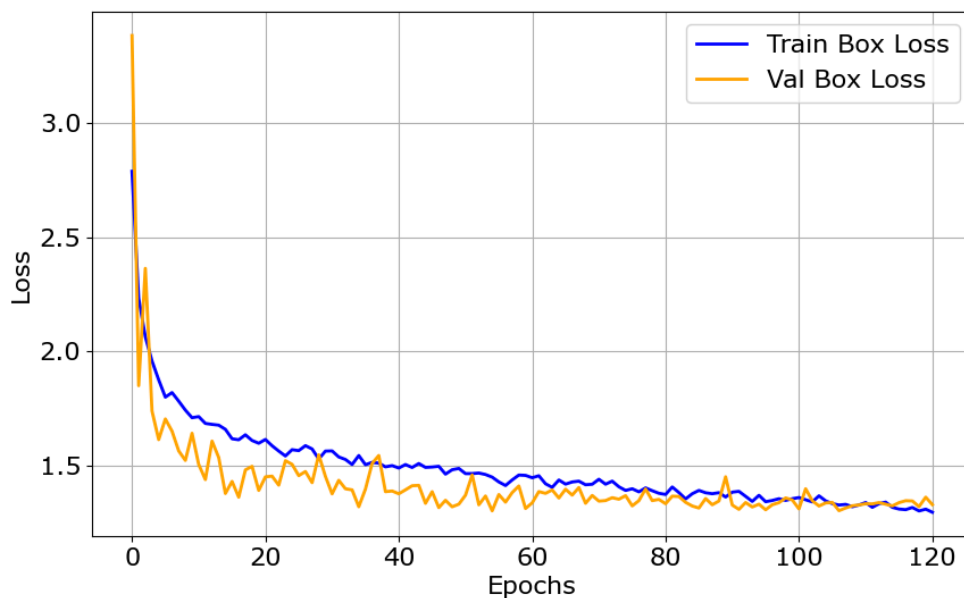


Figura 2. Pérdida de entrenamiento y validación.

Además, la Figura 2 presenta las curvas de pérdida de entrenamiento y validación, las cuales siguen trayectorias muy similares a lo largo de las épocas. Esta alineación confirma que los modelos lograron una buena generalización sin sobreajuste, ya que no se evidenció una brecha significativa entre el rendimiento en entrenamiento y validación.

Con base en los resultados mostrados en la Tabla 1, el modelo seleccionado como el mejor fue YOLOv11 con umbral 0.6, ya que proporciona mayor confianza en sus predicciones en comparación con el modelo pivote (YOLOv11 con umbral 0.5), manteniendo un rendimiento estadísticamente equivalente.

3.2. Evaluación del rendimiento en la fase de prueba

En la fase de evaluación y prueba de nuestra configuración experimental, evaluamos el rendimiento de los modelos YOLO entrenados utilizando un conjunto de prueba separado que no fue visto durante el entrenamiento ni la validación. Calculamos métricas clave de detección de objetos como $mAP@50$, precisión y recall para medir la exactitud y robustez de los modelos. La evaluación se realizó a través de diferentes umbrales de confianza (rango de 0.5 a 0.9) para observar la consistencia de los modelos bajo distintas exigencias de detección. Además, utilizamos pruebas estadísticas como la prueba de rangos con signo de Wilcoxon y la prueba U de Mann-Whitney para comparar el rendimiento de las diferentes versiones de los modelos YOLO, asegurando que las diferencias observadas fueron estadísticamente significativas y no debidas a variaciones aleatorias.

La evaluación del modelo seleccionado YOLOv11 en el conjunto de prueba externo confirmó su capacidad de generalización, alcanzando un $mAP@50$ de 0.868 (ver Tabla 1), con alta precisión (0.979) y un recall sólido (0.762) en un umbral de confianza de 0.6. Estos resultados demuestran una gran fiabilidad del modelo y un aprendizaje efectivo a partir del conjunto de datos de entrenamiento. Los resultados visuales de las detecciones en el conjunto de prueba, mostrados en la Figura 3, respaldan aún más estos hallazgos al ilustrar la colocación exitosa de los cuadros delimitadores en diversas muestras de mariquitas bajo diferentes condiciones (por ejemplo, iluminación variable, fondos naturales y oclusiones).

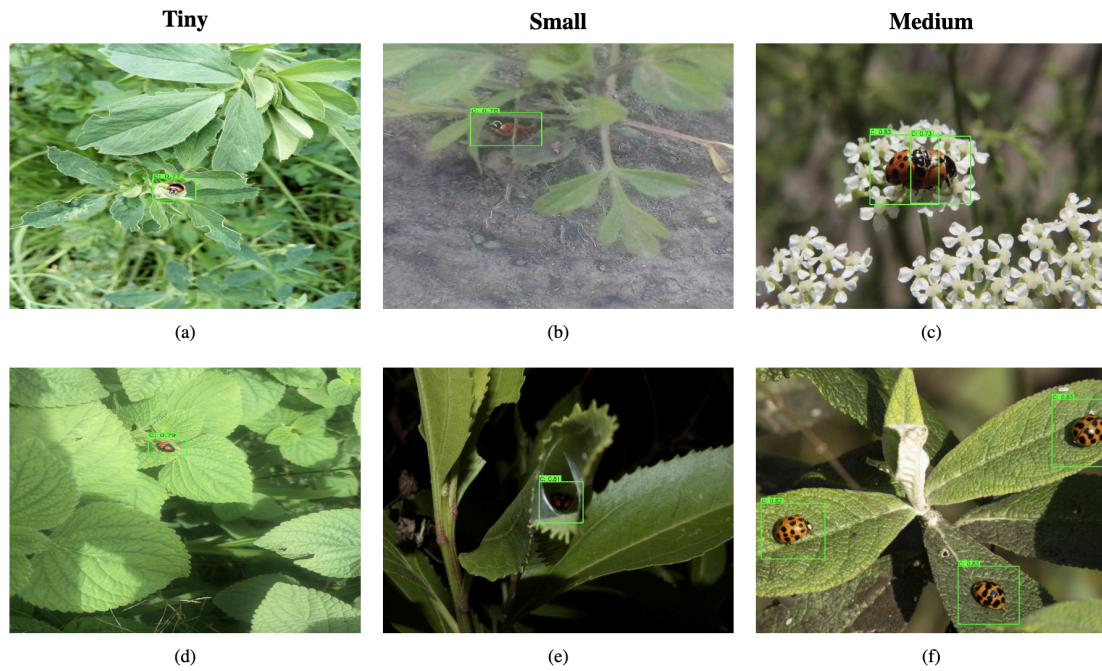


Figura 3. Comparación visual entre imágenes de mariquitas pequeñas, medianas y diminutas.

Sin embargo, la Figura 4 presenta ejemplos donde el modelo no logró un buen rendimiento. En muchos de estos casos fallidos, los problemas surgieron debido al tamaño pequeño del objeto, oclusiones intensas o patrones de fondo que imitaban las características de los insectos. Estos desafíos afectaron el recall del modelo de manera más significativa que la precisión, lo que significa que el modelo rara vez cometió falsas detecciones, pero a veces omitió verdaderos positivos. Este comportamiento es esperado cuando las pistas espaciales son débiles o cuando los insectos se desvían de los patrones dominantes aprendidos durante el entrenamiento.

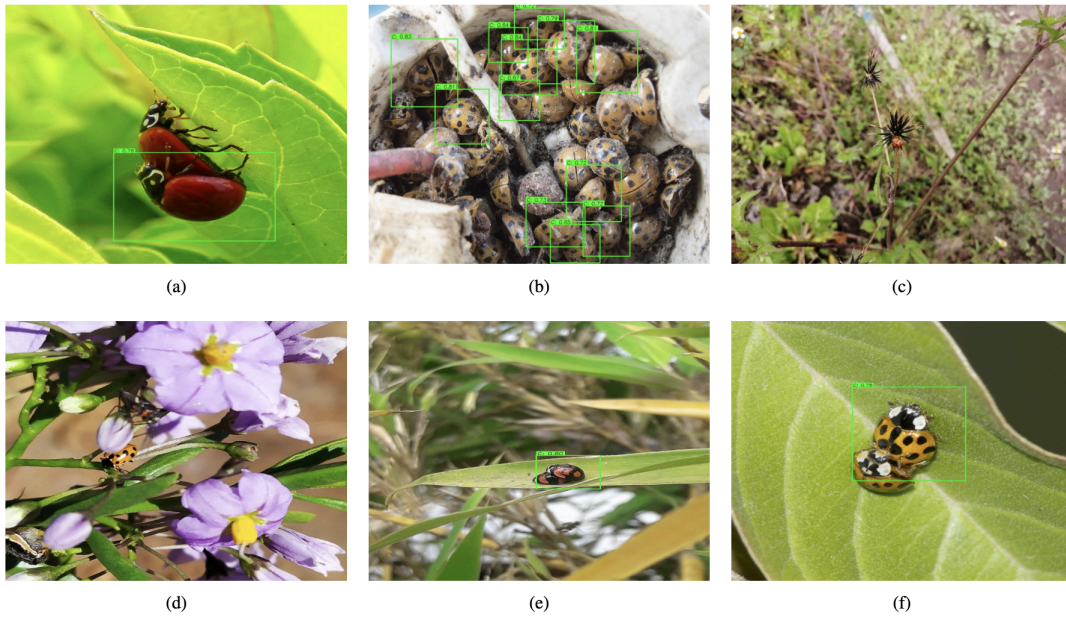


Figura 4. Escarabajos mariquitas que presentaron desafíos en la detección.

Al triangular los datos (alta variabilidad en la apariencia de los insectos y complejidad de las escenas), la arquitectura del modelo (YOLOv11 con mecanismos de atención y BiFPN), y los resultados de las métricas (descenso en el recall para ciertos escenarios), concluimos que, aunque el modelo es altamente preciso, su sensibilidad podría mejorarse en casos con bajo contraste visual o fondos recargados. Este hallazgo apunta hacia posibles mejoras en futuras iteraciones, como la incorporación de muestras de entrenamiento más diversas o el refinamiento de las capas de agregación de características.

3.3. Comparación con el estado del arte

Realizamos una comparación directa en la tarea de detección de mariquitas contra el modelo propuesto por Venegas et al. (2021), ya que su enfoque utiliza un subconjunto del mismo conjunto de datos utilizado en este estudio, enfocándose en el mismo dominio del problema. Su modelo alcanzó una precisión del 92% en una tarea de

clasificación. En contraste, nuestro detector basado en YOLOv11, diseñado para detección y localización de objetos, alcanzó una precisión de 0.979 y un recall de 0.762 en el conjunto de datos completo. Aunque las tareas difieren ligeramente —clasificación vs. detección de objetos—, la comparación destaca la robustez de nuestro modelo para manejar la localización espacial bajo condiciones de datos similares.

La comparación con otros enfoques de estado del arte es aproximada, ya que esos modelos fueron evaluados en conjuntos de datos y configuraciones de problema diferentes. Por ejemplo, YOLO-Pest (Xiang et al., 2023) reportó un $mAP@0.5$ de 91.9% en el conjunto de datos de plagas Teddy Cup, enfocándose en la detección de objetos pequeños en escenas agrícolas. Insect-YOLO (N. Wang et al., 2025), que integra mecanismos de atención para la detección de insectos en cultivos, alcanzó un $mAP@0.5$ de 93.8%. Aunque estos modelos logran un alto rendimiento en detección, sus resultados no son directamente comparables debido a diferencias en los conjuntos de datos, las especies de insectos objetivo y las condiciones ambientales. No obstante, nuestro modelo YOLOv11 alcanzó un sólido $mAP@0.5$ de 0.868 en el conjunto de prueba (umbral de confianza = 0.6), demostrando un rendimiento competitivo en el contexto específico de la detección de escarabajos mariquitas en escenas naturales complejas.

4. Conclusiones y Trabajo Futuro

Este trabajo presentó un enfoque de detección basado en YOLO para identificar escarabajos mariquitas en imágenes de la naturaleza. El modelo fue entrenado y evaluado en un conjunto de datos curado de 2899 imágenes de la plataforma iNaturalist Ciencia Ciudadana, cubriendo diversas regiones de Sudamérica. Entre las arquitecturas probadas, YOLOv11 alcanzó el mejor rendimiento, con un $mAP@50$ de 0.876 durante el entrenamiento y 0.868 en el conjunto de prueba, con un umbral de confianza de 0.6. Estos resultados demuestran la robustez y la capacidad de generalización del modelo en condiciones ecológicas reales. La efectividad de la pipeline propuesta resalta su potencial para apoyar el monitoreo automatizado de insectos y contribuir a investigaciones ecológicas y agrícolas más amplias.

Para trabajos futuros, planeamos expandir la pipeline de detección para que pueda soportar una mayor variedad de especies de insectos, utilizando bases de datos diversas de mariquitas y otros insectos. Esta expansión mejorará la capacidad del modelo para realizar monitoreo de biodiversidad a gran escala en diferentes ecosistemas. Al adaptar el modelo para detectar diversos grupos de insectos, podremos mejorar su utilidad tanto en investigación ecológica como en estrategias sostenibles de manejo de plagas.

Referencias bibliográficas

- Ahmad, I., Yang, Y., Yue, Y., Ye, C., Hassan, M., Cheng, X., ... Zhang, Y. (2022). Deep learning based detector YOLOv5 for identifying insect pests. *Applied Sciences*, 12(19), 10167. doi:10.3390/app121910167
- Alif, M. A. R., & Hussain, M. (2025). YOLOv12: A Breakdown of the Key Architectural Features. *arXiv Preprint arXiv:2502.14740*. doi:10.48550/arXiv.2502.14740
- Bhadane, G., Sharma, S., & Nerkar, V. B. (2013). Early pest identification in agricultural crops using image processing techniques. *International Journal of Electrical, Electronics and Computer Engineering*, 2(2), 77–82.
- Bjerger, K., Alison, J., Dyrmann, M., Frigaard, C. E., Mann, H. M. R., & Høye, T. T. (2023). Accurate detection and identification of insects from camera trap images with deep learning. *PLOS Sustainability and Transformation*, 2(3), e0000051. doi:10.1371/journal.pstr.0000051
- Cardim Ferreira Lima, M., Damascena de Almeida Leandro, M. E., Valero, C., Pereira Coronel, L. C., & Gonçalves Bazzo, C. O. (2020). Automatic detection and monitoring of insect pests—a review. *Agriculture*, 10(5), 161. doi:10.3390/agriculture10050161
- Deng, L., Wang, Y., Han, Z., & Yu, R. (2018). Research on insect pest image detection and recognition based on bio-inspired methods. *Biosystems Engineering*, 169, 139–148. doi:10.1016/j.biosystemseng.2018.02.008
- Dixon, A. F. G. (2000). *Insect predator-prey dynamics: ladybird beetles and biological control*. Cambridge University Press.
- Filali, I., Achour, B., Belkadi, M., & Lalam, M. (2022). Graph ranking based butterfly segmentation in ecological images. *Ecological Informatics*, 68, 101553. doi:10.1016/j.ecoinf.2022.101553
- Fu, X., Ma, Q., Yang, F., Zhang, C., Zhao, X., Chang, F., & Han, L. (2024). Crop pest image recognition based on the improved ViT method. *Information Processing in Agriculture*, 11(2), 249–259. doi:10.1016/j.inpa.2023.02.007
- Gondal, M. D., & Khan, Y. N. (2015). Early pest detection from crop using image processing and computational intelligence. *FAST-NU Research Journal*, 1(1), 59–68.
- Hidayatullah, P., Syakrani, N., Sholahuddin, M. R., Gelar, T., & Tubagus, R. (2025). YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review. *arXiv Preprint arXiv:2501.13400*. doi:10.48550/arXiv.2501.13400
- Huang, M.-L., Chuang, T.-C., & Liao, Y.-C. (2022). Application of transfer learning and image augmentation technology for tomato pest identification. *Sustainable*

- Computing: Informatics and Systems, 33, 100646.
doi:10.1016/j.suscom.2021.100646
- Kasinathan, T., Singaraju, D., & Uyyala, S. R. (2021). Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture*, 8(3), 446–457. doi:10.1016/j.inpa.2020.09.006
- Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv Preprint arXiv:2410.17725*. doi:10.48550/arXiv.2410.17725
- Khujamatov, H., Muksimova, S., Abdullaev, M., Cho, J., & Jeon, H.-S. (2025). Advanced Insect Detection Network for UAV-Based Biodiversity Monitoring. *Remote Sensing*, 17(6), 962. doi:'10.3390/rs17060962'
- Kittichai, V., Pengsakul, T., Chumchuen, K., Samung, Y., Sriwichai, P., Phatthamolrat, N., ... Boonsang, S. (2021). Deep learning approaches for challenging species and gender identification of mosquito vectors. *Scientific Reports*, 11(1), 4838. doi:10.1038/s41598-021-84219-4
- Mamdouh, N., & Khattab, A. (2021). YOLO-Based Deep Learning Framework for Olive Fruit Fly Detection and Counting. *IEEE Access*, 9, 84252–84262. doi:10.1109/ACCESS.2021.3088075
- Nanni, L., Maguolo, G., & Pancino, F. (2020). Insect pest image detection and recognition based on bio-inspired methods. *Ecological Informatics*, 57, 101089. doi:10.1016/j.ecoinf.2020.101089
- Nanni, L., Manfè, A., Maguolo, G., Lumini, A., & Brahnam, S. (2022). High performing ensemble of convolutional neural networks for insect pest image detection. *Ecological Informatics*, 67, 101515. doi:10.1016/j.ecoinf.2021.101515
- Simberloff, D., & Stiling, P. (1996). Risks of species introduced for biological control. *Biological Conservation*, 78(1–2), 185–192. doi:10.1016/0006-3207(96)00027-4
- Teixeira, A. C., Ribeiro, J., Morais, R., Sousa, J. J., & Cunha, A. (2023). A systematic review on automatic insect detection using deep learning. *Agriculture*, 13(3), 713. doi:'10.3390/agriculture13030713'
- Ultralytics. (2023). Performance Metrics Deep Dive. Retrieved 12 May 2025, from <https://docs.ultralytics.com/guides/yolo-performance-metrics/>
- Ultralytics. (2025). Ultralytics YOLO11. Retrieved from <https://github.com/ultralytics/ultralytics>
- Venegas, P., Calderon, F., Riofrío, D., Benítez, D., Ramón, G., Cisneros-Heredia, D., & Others. (2021). Automatic ladybird beetle detection using deep-learning models. *PLoS ONE*, 16(6). doi:10.1371/journal.pone.0253027

- Walter, T., Degen, J., Pfeiffer, K., Stöckl, A., Montenegro, S., & Degen, T. (2021). A new innovative real-time tracking method for flying insects applicable under natural conditions. *BMC Zoology*, 6, 1–11.
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Others. (2024). YOLOv10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems*, 37, 107984–108011. doi:10.48550/arXiv.2405.14458
- Wang, J., Hong, M., Hu, X., Li, X., Huang, S., Wang, R., & Zhang, F. (2023). Camouflaged insect segmentation using a progressive refinement network. *Electronics*, 12(4), 804. doi:10.3390/electronics12040804
- Wang, N., Fu, S., Rao, Q., Zhang, G., & Ding, M. (2025). Insect-YOLO: A new method of crop insect detection. *Computers and Electronics in Agriculture*, 232, 110085. doi:10.1016/j.compag.2025.110085
- Wang, X., Zhang, S., Wang, X., & Xu, C. (06 2023). Crop pest detection by three-scale convolutional neural network with attention. *PLOS ONE*, 18(6), 1–16. doi:10.1371/journal.pone.0276456
- Wang, Z., Wang, K., Liu, Z., Wang, X., & Pan, S. (2018). A cognitive vision method for insect pest image segmentation. *IFAC-PapersOnLine*, 51(17), 85–89. doi:10.1016/j.ifacol.2018.08.066
- Xia, D., Chen, P., Wang, B., Zhang, J., & Xie, C. (2018). Insect detection and classification based on an improved convolutional neural network. *Sensors*, 18(12), 4169. doi:10.3390/s18124169
- Xiang, Q., Huang, X., Huang, Z., Chen, X., Cheng, J., & Tang, X. (2023). YOLO-pest: An insect pest object detection algorithm via CAC3 module. *Sensors*, 23(6), 3221. doi:10.3390/s23063221
- Xiong, P., Zhang, C., He, L., Zhan, X., & Han, Y. (11 2024). Deep learning-based rice pest detection research. *PLOS ONE*, 19(11), 1–20. doi:10.1371/journal.pone.0313387