

Arquitecturas YOLO y RF DETR

para el diagnóstico temprano de enfermedades en
Solanum tuberosum L

Manolo Sebastián Muñoz-Espinoza¹

¹Universidad Técnica de Ambato, Facultad
de Ciencias Agropecuarias, Ambato,
Ecuador,
<https://orcid.org/0000-0001-5897-1211>
mmunoz@uta.edu.ec

DOI: 10.61854/rccedia.v1n1.005

Enviado: 12 de febrero de 2026

Aprobado: 06 de abril de 2026

Publicado: martes 14 de julio de 2026



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0.

Resumen

Se compararon, bajo un protocolo común, tres detectores de objetos para el diagnóstico poscosecha en tubérculos: dos arquitecturas de la familia YOLO y un detector basado en transformadores (RF-DETR). Se empleó un conjunto de datos público de seis clases (cinco patologías y una sana). Las lesiones se anotaron con bounding boxes en Roboflow y se aplicó aumento de datos geométrico y fotométrico. Tras la conversión a COCO para RF-DETR, se mantuvieron idénticas particiones (train/val/test = 792/56/57), misma resolución (640 px) y el mismo protocolo de entrenamiento (60 épocas con early stopping). La evaluación se realizó exclusivamente en test con precisión, recall, mAP@0.50 y mAP@0.50:0.95. En test, YOLOv12 alcanzó P = 78,72 %, R = 96,86 %, mAP@0.50 = 79,36 %, mAP@0.50:0.95 = 73,26 %; YOLOv26 obtuvo P = 79,34 %, R = 96,64 %, mAP@0.50 = 78,48 %, mAP@0.50:0.95 = 73,55 % (según los registros experimentales). RF-DETR logró P = 70,86 %, R = 88,39 %, mAP@0.50 = 85,61 % y mAP@0.50:0.95 = 82,64 %. Para el tamizado de alta sensibilidad (recall alto), la arquitectura YOLO es más conveniente, mientras que RF-DETR ofrece localización estricta (mAP superior en 0.50–0.95) y simplifica el posprocesamiento al prescindir de NMS. Un pipeline en dos etapas (YOLO → RF-DETR) combina velocidad y precisión espacial. Como trabajo futuro, se sugiere calibración por clase y fusión con NIR/HSI para lesiones subepidérmicas.

Palabras clave:

agricultura de precisión, aprendizaje profundo, aumento de datos, localización de objetos, visión por computadora.

Introducción

La papa (*Solanum tuberosum* L.) es uno de los cultivos alimentarios relevantes del mundo y ocupa el tercer lugar en consumo global (Nagel y otros, 2022). En el 2022, la producción superó 375 millones de toneladas, con China e India como principales productores (95,5 y 56 millones de toneladas, respectivamente) (Guillemette y otros, 2024; Solankey, 2025). Su relevancia se sustenta no solo en su valor nutricional y versatilidad culinaria, sino también en su papel estratégico para la seguridad alimentaria y la economía rural, especialmente en países en desarrollo (Moreno Mayhuire & Herrera Quispe, 2023). Sin embargo, la productividad y la calidad poscosecha se ven afectadas por enfermedades fúngicas, bacterianas y fisiológicas que impactan directamente los tubérculos (Keithellakpam y otros, 2025).

Entre las patologías de relevancia económica se encuentran la pierna negra (*Pectobacterium* spp.), la podredumbre rosada (*Phytophthora erythroseptica*), la pudrición seca (*Fusarium* spp.), la sarna común (*Streptomyces scabies*) y la sarna negra (*Rhizoctonia solani*) (Sharma y otros, 2024; Upeniece y otros, 2025). Estas enfermedades pueden ocasionar pérdidas poscosecha del 30 % al 50 % cuando no se detectan ni gestionan oportunamente (Aleinzi, 2025).

El diagnóstico temprano es un componente necesario en la cadena de valor agrícola. No obstante, los métodos basados en la inspección visual experta suelen ser subjetivos, lentos, costosos y dependientes de la experiencia humana, lo cual limita su escalabilidad y reproducibilidad (Qin y otros, 2024). En contraste, los enfoques de inteligencia artificial (IA), en particular el aprendizaje profundo aplicado a la visión por computadora, han emergido como herramientas útiles en la agricultura de precisión (Jayanthi & Kumar, 2024). Los

modelos de detección de objetos ofrecen ventajas frente a los enfoques de clasificación al permitir no solo identificar la presencia de una enfermedad, sino también localizarla en la superficie del tubérculo (Ma y otros, 2022).

En este contexto, Ultralytics YOLO (You Only Look Once) ha demostrado un equilibrio entre precisión, velocidad y eficiencia computacional, lo que ha motivado su adopción en aplicaciones agrícolas (Divyanth y otros, 2024; Geoffrey Manurung y otros, 2024; LIU, 2024; Palacio Betancur & Bolaños Martínez, 2025; Redmon, 2016). Versiones previas, como YOLOv5, YOLOv8 y YOLOv12, han mostrado resultados prometedores en la detección de plagas y enfermedades en cultivos como tomate, fresa y papa (He y otros, 2024; Manolo Muñoz-Espinoza y otros, 2026; J. Zhang y otros, 2024).

YOLO26 es una versión reciente para detección de objetos en tiempo real, diseñada para equipos de borde y con bajo consumo energético. Emplea un diseño optimizado que reduce la complejidad computacional, lo que facilita una implementación ágil, liviana y accesible (Jocher & Qiu, 2026; Sapkota y otros, 2025; Yurdakul y otros, 2026).

Del mismo modo, RF-DETR es una arquitectura de detección de objetos en tiempo real basada en transformadores, perteneciente a la familia de los DETR (*detection transformers*). Se ajusta a diversos dominios y conjuntos de datos de distintos tamaños y está orientada a aplicaciones que requieren alta velocidad y precisión, incluso en entornos con recursos computacionales limitados (por ejemplo, dispositivos de borde o aplicaciones de baja latencia) (Robinson y otros, 2026).

A pesar de estos avances, persisten desafíos: (1) la escasez de conjuntos de datos locales, públicos y etiquetados que representen condiciones reales de campo o poscosecha; (2) la necesidad de adaptar modelos a contextos con muestras limitadas mediante un aumento de datos robusto; y (3) la validación rigurosa en tareas de detección multiclase con clases visualmente similares.

Con el objetivo de comparar prioridades de diseño complementarias, en este estudio se analizaron dos detectores representativos de la familia YOLO —que priorizan velocidad e implementación sencilla— y se amplió el alcance incorporando un detector basado en transformadores, RF-DETR, que formula la detección como predicción de conjuntos a partir de consultas (*queries*) y una asignación uno a uno, reduciendo la dependencia de anclajes y el posprocesamiento intensivo (Robinson y otros, 2026). Para asegurar la equidad experimental, los tres detectores se entrenaron con las mismas imágenes, idénticas particiones (entrenamiento/validación/prueba) y la misma resolución de entrada, manteniendo un protocolo de evaluación unificado. Las anotaciones se convirtieron al formato COCO para el entrenamiento del modelo de tipo DETR.

El enfoque propuesto emplea un conjunto de datos abierto de tubérculos de papa (Faria y otros, 2023), con seis clases (cinco enfermedades y una condición sana), anotado en Roboflow (*Roboflow: Computer Vision Tools for Developers and Enterprises*, 2025) y ampliado mediante un protocolo de aumento diseñado para simular variabilidad realista en orientación, iluminación y ruido.

El objetivo principal fue evaluar comparativamente el rendimiento de detectores de la familia

YOLO y detectores basados en transformadores (RF-DETR) en el descubrimiento temprano de enfermedades en tubérculos de papa, con miras a desarrollar una herramienta automatizada, precisa y escalable que pueda integrarse en sistemas de monitoreo en tiempo real y en entornos con recursos limitados.

Contribuciones

- Un *benchmark* comparativo entre detectores *CNN-based* de la familia YOLO y un detector *transformer-based* (RF-DETR), utilizando imágenes agrícolas reales de papa con seis clases, caracterizadas por alta variabilidad visual y condiciones de adquisición no controladas.
- Un *pipeline* reproducible de anotación y conversión a formato COCO, con criterios de comparabilidad definidos, incluyendo particiones idénticas, resolución uniforme y estrategias de aumento consistentes.
- Una evaluación comparativa homogénea y orientada al análisis del comportamiento de los modelos, que complementa las métricas estándar (mAP@0.5 y mAP@0.5:0.95) con métricas de *precision*, *recall* y matrices de confusión, permitiendo identificar errores sistemáticos y desafíos específicos del dominio agrícola.

Métodos

El estudio presenta un diseño experimental comparativo de tres detectores de objetos (YOLOv12, YOLOv26 y RF-DETR) bajo un protocolo común. La secuencia completa del procesamiento fue: (1) obtención y etiquetado manual de lesiones en 377 imágenes originales; (2) división estratificada en entrenamiento (70 %), validación (15 %) y prueba (15 %); (3) aumento exclusivo del conjunto de entrenamiento; (4) conversión de anotaciones a formato COCO para RF-DETR; (5) entrenamiento independiente de cada modelo con

igual resolución (640 px), mismas particiones y criterios de evaluación; (6) evaluación final sobre el conjunto de prueba con métricas estandarizadas (*precision*, *recall*, $mAP@0.50$, $mAP@0.50:0.95$).

Conjunto de datos original y criterios de inclusión/exclusión

El conjunto de datos base se obtuvo de la plataforma Kaggle (Faria y otros, 2023).

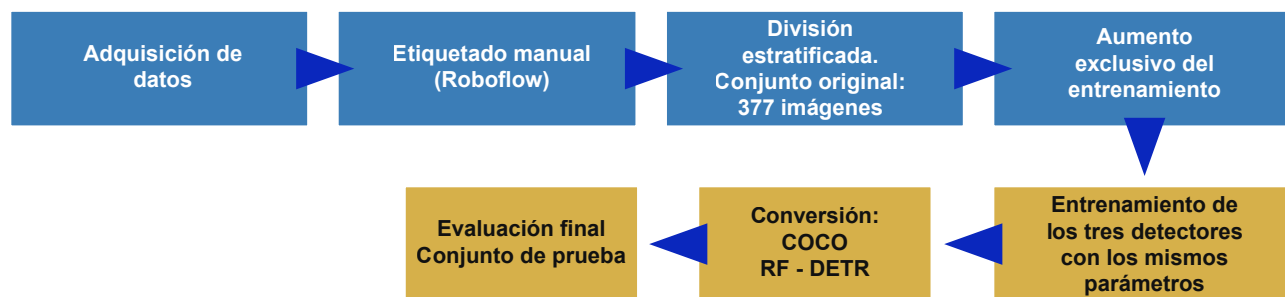


Figura 1

Secuencia completa del pipeline experimental, desde la adquisición de datos hasta la evaluación comparativa

Nota. Las etapas incluyen: etiquetado manual en Roboflow, división estratificada del conjunto original (377 imágenes), aumento exclusivo del entrenamiento, conversión a COCO para RF-DETR, entrenamiento de los tres detectores con mismos parámetros base (640 px, mismas particiones) y evaluación final en el conjunto de prueba.

El conjunto de datos se compiló a partir de múltiples fuentes en línea y fue validado por el Instituto de Investigación Agrícola de Bangladesh (BARI). Consta de 451 imágenes de tubérculos de papa. En este estudio se utilizaron 377 imágenes, excluyendo la clase “misceláneos” para mantener un enfoque estrictamente fitosanitario. Las seis clases consideradas y su distribución fueron:

- Pierna negra: 60 imágenes
- Podredumbre rosada: 57
- Pudrición seca: 60
- Sarna común: 62
- Sarna negra: 58
- Sana: 80

Etiquetado de objetos y control de calidad

El etiquetado manual de las lesiones se realizó con Roboflow (*Roboflow: Computer Vision Tools for Developers and Enterprises, 2025*). Cada instancia de lesión se definió mediante un rectángulo delimitador (*bounding box*), procurando cubrir la lesión visible sin incluir áreas sanas adyacentes en exceso. Para controlar la calidad de las anotaciones:

- Los ejemplos de cada clase fueron revisados visualmente por un especialista en fitopatología, con el fin de verificar la coherencia en el criterio de las anotaciones.
- Se corrigieron cajas excesivamente holgadas o que no coincidían con el contorno aparente de la lesión.
- Las anotaciones se gestionaron en los formatos YOLO y COCO, generados a partir de un mismo conjunto base, con el fin de asegurar la compatibilidad con los frameworks de entrenamiento y evaluación utilizados.

Preprocesamiento

Antes del entrenamiento, se aplicaron las transformaciones:

- Orientación automática para normalizar la rotación del objeto principal (tubérculo).
- Redimensionamiento de las imágenes a 640 × 640 px, manteniendo la relación de aspecto mediante *letterboxing* cuando fue necesario (comportamiento por defecto de Ultralytics).

División estratificada de datos

A partir del conjunto original de 377 imágenes (sin aumento de datos), se realizó una división estratificada para preservar la proporción de clases en cada subconjunto:

- Entrenamiento: 264 imágenes (70 %).
- Validación: 56 imágenes (15 %).
- Prueba: 57 imágenes (15 %).

Aumento de datos (aplicado solo al conjunto de entrenamiento)

Con el fin de incrementar la diversidad de patrones y mitigar el sobreajuste, se aplicó un *pipeline* de aumento exclusivamente sobre las 264 imágenes del conjunto de entrenamiento. Se utilizaron transformaciones en Roboflow que preservan la coherencia geométrica entre imagen y *bounding boxes*:

Transformaciones de imagen:

- Volteo horizontal (*horizontal flip*).
- Rotaciones rígidas de 90° (horaria y antihoraria).
- Rotación aleatoria en el rango [-15° a +15°].
- Inyección de ruido aleatorio hasta el 0,65 % de los píxeles.

Transformaciones de *bounding box*:

- Volteo horizontal sincronizado con la imagen.
- Rotación en $[-15^\circ \text{ a } +15^\circ]$.
- Cizallamiento (*shear*) horizontal y vertical de $\pm 10^\circ$.
- Ruido posicional limitado (hasta 0,16 % de los píxeles) para robustecer la sensibilidad a pequeñas variaciones de anotación.

Este proceso produjo un conjunto de entrenamiento aumentado de 792 imágenes (factor de aumento $\sim 3\times$). Los conjuntos de validación (56 imágenes) y prueba (57 imágenes) se mantuvieron sin aumento.

- Entrenamiento: 792 imágenes ($\sim 88\%$).
- Validación: 56 imágenes ($\sim 6\%$).
- Prueba: 57 imágenes ($\sim 6\%$).

Estos porcentajes se calculan sobre el total de imágenes utilizadas en el experimento ($905 = 792 + 56 + 57$).

La división estratificada $\sim 88\% / \sim 6\% / \sim 6\%$ se seleccionó teniendo en cuenta las prácticas comunes en tareas de detección de objetos con redes profundas, en las que se privilegia un conjunto de entrenamiento amplio para favorecer la convergencia y la capacidad de generalización del modelo. Dado el tamaño limitado del conjunto de datos, se optó por un conjunto de validación reducido pero representativo, suficiente para el monitoreo del entrenamiento y el ajuste de hiperparámetros, y por un conjunto de prueba independiente para la evaluación final.

La estratificación asegura la preservación de la proporción de clases en todos los subconjuntos, reduciendo sesgos derivados del desbalance de clases y permitiendo una comparación consistente del desempeño entre modelos.

Modelos de detección: YOLOv12 y YOLOv26

- YOLOv12 (nano) - *yolov12n.pt*. Detector reciente centrado en mecanismos de atención para enfatizar características relevantes y mejorar la detección de objetos pequeños o parcialmente ocluidos. El modelo se obtuvo del repositorio oficial (Tian y otros, 2025).
- YOLOv26 (nano) - *yolov26n.pt*. Variante optimizada para dispositivos periféricos (*edge*) con recursos limitados; prioriza simplicidad arquitectónica y eficiencia en inferencia (Jocher & Qiu, 2026).
- RF-DETR (base). Arquitectura de transformador en tiempo real para la detección de objetos y la segmentación de instancias, desarrollada por Roboflow (Robinson y otros, 2026).

Para garantizar la comparabilidad, los modelos se entrenaron y evaluaron con el mismo conjunto de datos, idéntica partición (train/val/test), misma resolución de entrada (640×640 px) y mismo protocolo de entrenamiento y evaluación.

Entorno de software y hardware

Los experimentos se ejecutaron en Google Colab con GPU NVIDIA Tesla T4 (parámetros en la Figura 2). Todo el software utilizado es de código abierto y gratuito: Ultralytics YOLO (licencia AGPL-3.0), RF-DETR (licencia Apache 2.0) y Google Colab (versión gratuita). Las anotaciones se realizaron con Roboflow (versión gratuita, licencia de uso académico).

NVIDIA-SMI 550.54.15				Driver Version: 550.54.15				CUDA Version: 12.4			
GPU	Name			Persistence-M	Bus-Id	Disp.A	Volatile	Uncorr.	ECC		
Fan	Temp	Perf		Pwr:Usage/Cap		Memory-Usage	GPU-Util	Compute	M.		
										MIG	M.
0	Tesla T4			Off	00000000:00:04.0	Off				0	
N/A	35C	P8		9W / 70W	0MiB / 15360MiB		0%	Default		N/A	

Processes:											
GPU	GI	CI	PID	Type	Process name					GPU Memory	
	ID	ID								Usage	
No running processes found											

Figura 2

Parámetros de Google Colab para el entrenamiento

Protocolo de entrenamiento

Pesos iniciales: Se usaron pesos preentrenados (*pretrained*) de cada variante “nano” correspondiente (yolov12n.pt y yolov26n.pt) y “base” de RF-DETR.

Para la arquitectura YOLO, en los dos modelos, los hiperparámetros fueron:

- Épocas (*epochs*): 60
- Tamaño de imagen (*imgsz*): 640 px
- Tamaño de lote (*batch*): 16
- Patience (*early_stopping_patience*): 10 (si no mejora la métrica de validación durante 10 épocas consecutivas)
- Archivo de configuración (*data.yaml*): rutas de train/val/test y definición de 6 clases
- Aumentaciones: las descritas en la sección de “División estratificada de datos” (aplicadas de manera consistente)

Para la arquitectura RF-DETR, los hiperparámetros fueron:

- Épocas (*epochs*): 15
- Tamaño de imagen (*imgsz*): 640 px
- Tamaño de lote (*batch*): 4
- Patience (*early_stopping_patience*): 5

Las diferencias en el tamaño del lote y número de épocas responden a las características propias de cada arquitectura y a restricciones de memoria, práctica común en estudios comparativos cuando se emplean *frameworks* distintos, manteniendo constantes el conjunto de datos, la resolución y los criterios de evaluación.

Evaluación y métricas

La evaluación se realizó exclusivamente sobre el conjunto de prueba (57 imágenes); se reportaron las métricas estándar de Ultralytics para tareas de detección:

- $mAP@0.5$: media del *average precision* por clase empleando un umbral de $IoU = 0.5$.

$$(3) \ mAP@0.50 = \frac{1}{N} \sum_{i=1}^N AP_i(IoU \geq 0.50)$$

- $mAP@0.5:0.95$: media del AP en múltiples umbrales de IoU [0.5: 0.95] con paso de 0.05, más exigente y sensible a la localización.

$$(4) \ mAP@0.50:0.95 = \frac{1}{10} \sum_{t=0.50}^{0.95} mAP_t$$

- $F1$ -score: combina la *precision* y el *recall*.

$$(5) \ F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Resultados

Métricas de rendimiento global

Los tres modelos se evaluaron en el conjunto de prueba (n = 57 imágenes), bajo las mismas condiciones. En la Figura 3 se observan las de-

tecciones de cada modelo y la Tabla 1 resume los resultados globales.

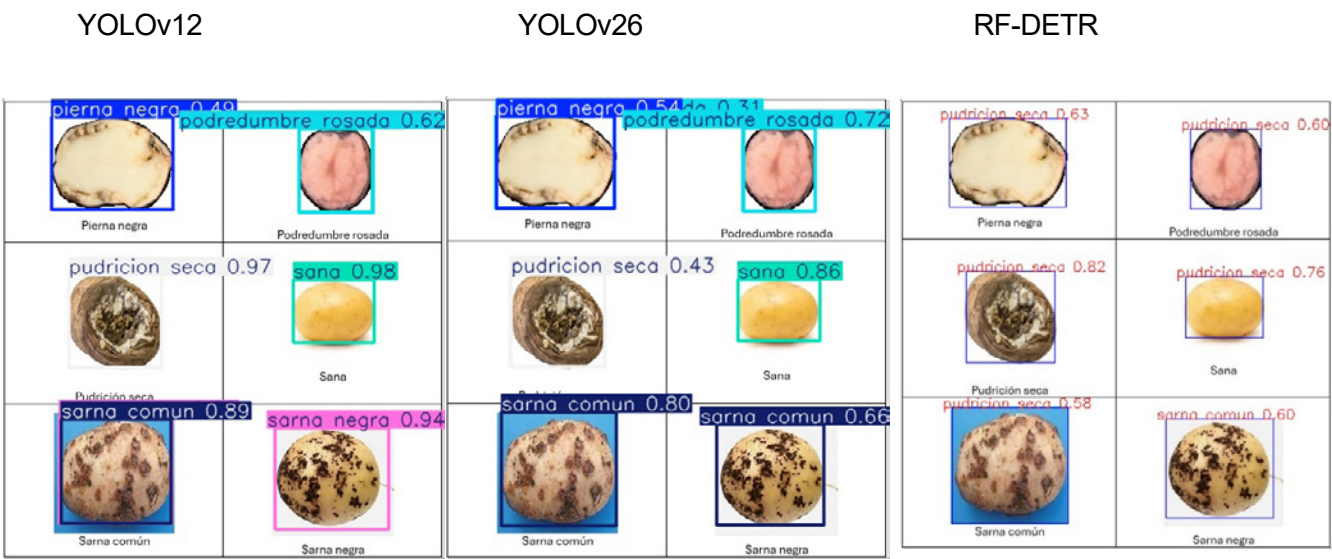


Figura 3
Comparación visual de los diferentes modelos

Tabla 1
Resultados globales

Modelo	<i>Precision</i>	<i>Recall</i>	mAP50	mAP95
RF-DETR	70,86 %	88,39 %	85,61 %	82,64 %
YOLOv1 2	78,72 %	96,86 %	79,36 %	73,26 %
YOLOv2 6	79,34 %	96,64 %	78,48 %	73,55 %

Matriz de confusión

La Figura 4 (YOLOv12) y la Figura 5 (YOLOv26) muestran las matrices de confusión obtenidas; en el eje diagonal se observan los aciertos por clase. Además de las seis clases del estudio (pierna negra, podredumbre rosada, pudrición seca, sarna común, sarna negra, sana), ambas matrices incluyen una columna/fila de *background*, que agrega:

- Falsos positivos (detecciones en regiones sin lesión asignable).
- Falsos negativos (lesiones omitidas que se contabilizan contra la clase esperada como *background*).

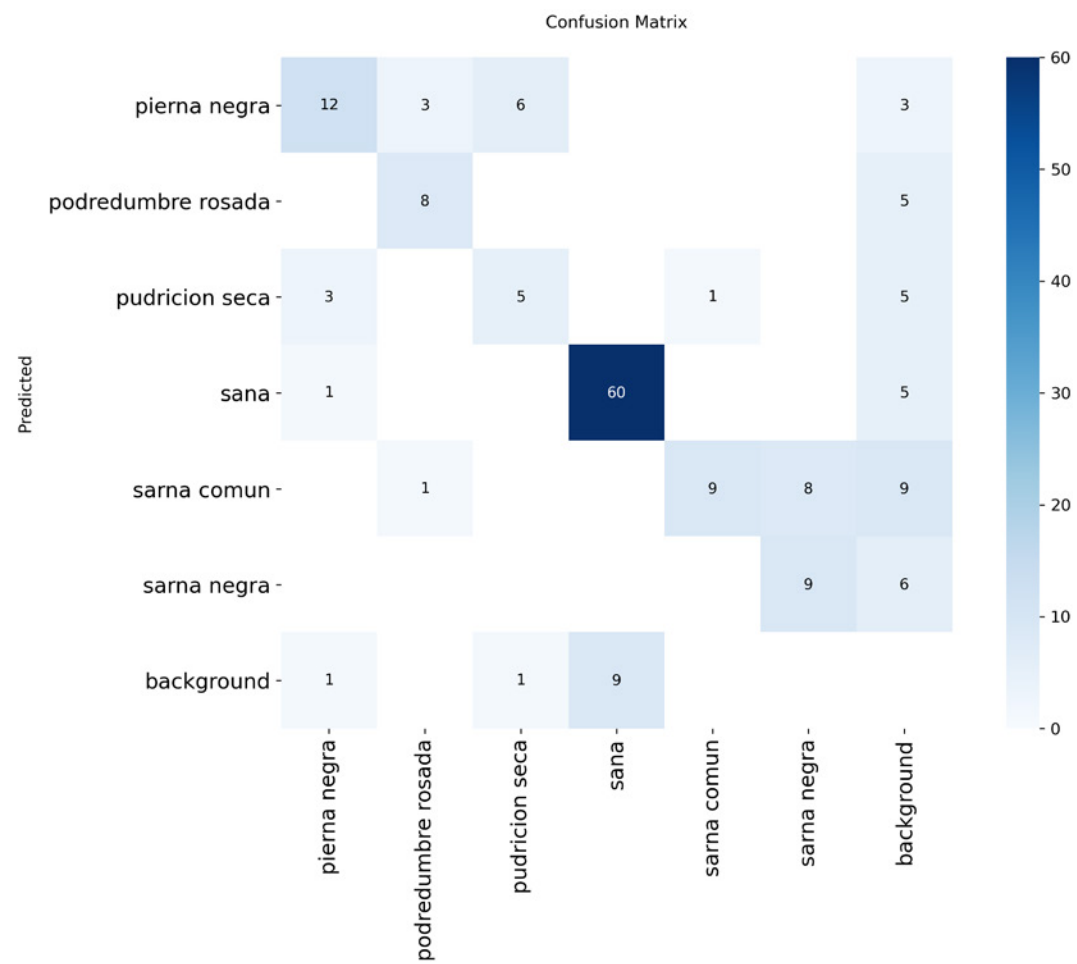


Figura 4
Matriz de confusión – YOLOv12

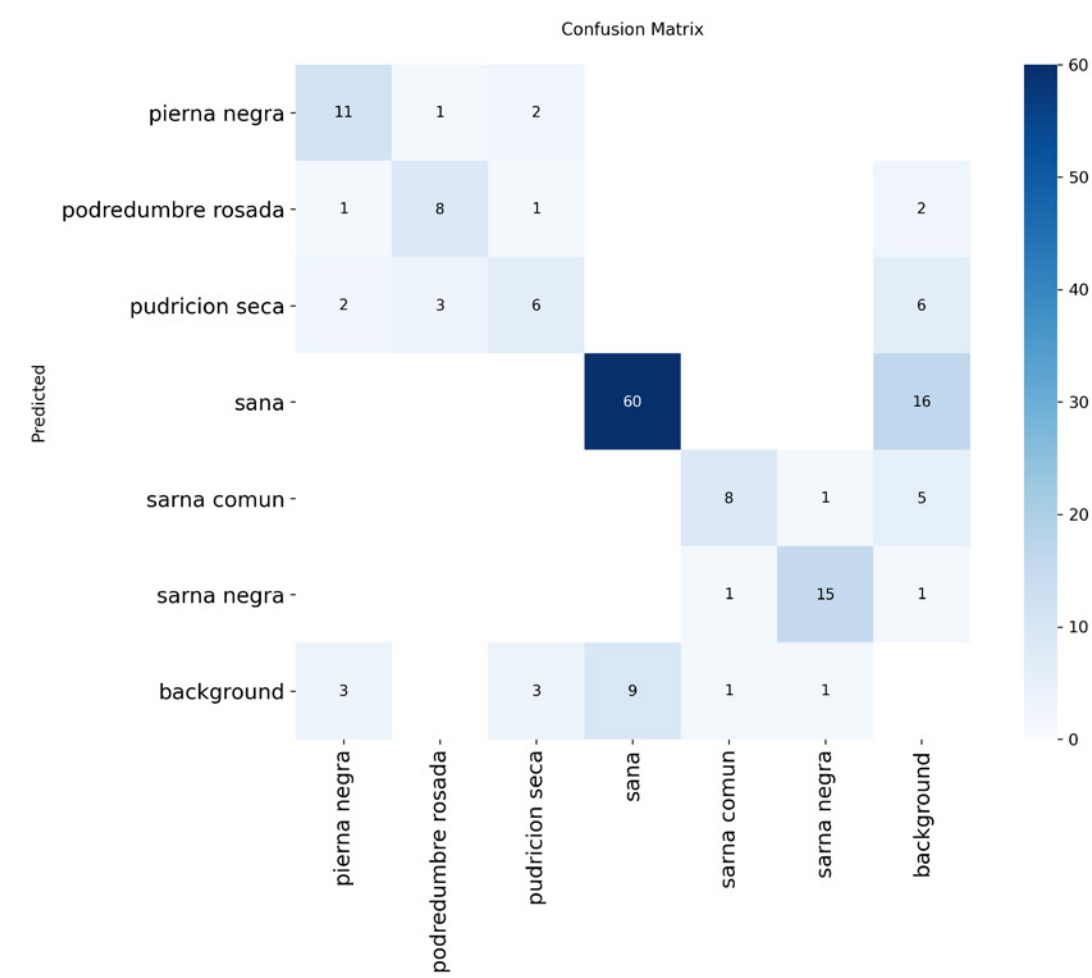


Figura 5
Matriz de confusión – YOLOv26

Desempeño por clase

A partir de las matrices de confusión, se calcularon: *precision* y *recall* por clase.

Tabla 2
Métricas por clase – YOLOv12

Clase	Precision	Recall
Pierna negra	50,00 %	75,00 %
Podredumbre rosada	61,54 %	66,67 %
Pudrición seca	35,71 %	41,67 %
Sarna común	33,33 %	90,00 %
Sarna negra	60,00 %	52,94 %
Sana	92,31 %	86,96 %

Tabla 3
Métricas por clase - YOLOv26

Clase	<i>Precision</i>	<i>Recall</i>
Pierna negra	78,57 %	64,71 %
Podredumbre rosada	61,54 %	66,67 %
Pudrición seca	35,71 %	41,67 %
Sarna común	33,33 %	90,00 %
Sarna negra	60,00 %	52,94 %
Sana	92,31 %	86,96 %

Tabla 4
Métricas por clase - RF-DETR

Clase	Precision	Recall
Pierna negra	67,00 %	35,00 %
Podredumbre rosada	100,00 %	75,00 %
Pudrición seca	53,00 %	75,00 %
Sarna común	50,00 %	100,00 %
Sarna negra	93,00 %	82,00 %
Sana	92,00 %	88,00 %

Hallazgos clave: comparación integral entre modelos

De las tablas 1, 2, 3 y 4, se obtienen hallazgos importantes que permiten realizar una comparación integral entre modelos.

3.4.1. Comparación global (exactitud y localización)

A partir de los resultados (Tabla 1), se observa:

- *Precision* global (P): YOLOv26 alcanza el valor más alto (79,34 %), seguido de YOLOv12 (78,72 %) y RF-DETR (70,86 %).
- *Recall* global (R): YOLOv12 es ligeramente superior (96,86 %), cerca de YOLOv26 (96,64 %) y por encima de RF-DETR (88,39 %).
- $mAP@0.50$: RF-DETR es el mejor (85,61 %) y supera a YOLOv12 (79,36 %) y YOLOv26 (78,48 %).
- $mAP@0.50:0.95$ (más estricto): RF-DETR también lidera (82,64 %), con YOLOv26 (73,55 %) y YOLOv12 (73,26 %) cercanos entre sí.

Los modelos YOLO presentan mayor precisión puntual y sensibilidad globales al umbral por defecto (alta capacidad para no omitir lesiones y controlar falsos positivos a un punto de operación fijo).

RF-DETR exhibe AP superior tanto a IoU laxo (0.50) como a IoU estricto (0.50–0.95), lo que sugiere que a lo largo del barrido de umbrales el modelo mantiene mejor el área bajo la curva *precision–recall* y produce cajas más consistentes en términos de localización.

Implicación práctica: Si la aplicación exige máxima sensibilidad y bajo costo de

omisión (tamizado), uno de los YOLO es una opción adecuada con el umbral por defecto. Si la prioridad es la calidad de localización (p. ej., estimar severidad por área, recortar lesiones para segmentar o integrar a un *pipeline* posterior), RF-DETR ofrece una ventaja consistente en mAP, especialmente en el rango de 0.50–0.95.

3.4.2. Comparación por clase (*precision*, *recall* y *F1*)

Resultados por clase

- Pierna negra
- YOLOv26: P = 78,57 % (máxima), R = 64,71 %, F1 $\approx$ 71,00 % (mejor balance).
- YOLOv12: P = 50,00 %, R = 75,00 % (máximo *recall*), F1 $\approx$ 60,00 %.
- RF-DETR: P = 67,00 %, R = 35,00 %, F1 $\approx$ 46,00 %.
- Conclusión: YOLOv26 equilibra mejor la clase; YOLOv12 recupera más verdaderos positivos (mayor R), útil para no omitir casos; RF-DETR necesita calibración del umbral o más muestras.

Podredumbre rosada

- RF-DETR: P = 100,00 %, R = 75,00 %, F1 $\approx$ 86,00 % (superior).
- YOLOv12/YOLOv26: P = 61,54 %, R = 66,67 %, F1 $\approx$ 64,00 %.
- Conclusión: RF-DETR es superior; combina cero falsos positivos con alta sensibilidad en esta clase.

Pudrición seca

- RF-DETR: P = 53,00 %, R = 75,00 %, F1 $\approx$ 62,00 % (mejor).
- YOLOv12/YOLOv26: P = 35,71 %, R = 41,67 %, F1 $\approx$ 39,00 %.
- Conclusión: La variabilidad morfológica de esta lesión (bordes difusos/subepidérmicos) perjudica a los YOLO con cajas; RF-DETR maneja mejor la localización y recupera más casos.

Sarna común

- RF-DETR: P = 50,00 %, R = 100,00 %, F1 $\approx$ 67,00 % (mejor).
- YOLOv12/YOLOv26: P = 33,33 %, R = 90,00 %, F1 $\approx$ 48,00 %.
- Conclusión: RF-DETR reduce la sobre-detección típica de esta clase en los YOLO (P baja) manteniendo *recall* máximo.

Sarna negra

- RF-DETR: P = 93,00 %, R = 82,00 %, F1 $\approx$ 88,00 % (muy superior).
- YOLOv12/YOLOv26: P = 60,00 %, R = 53,00 %, F1 $\approx$ 56,00 %.
- Conclusión: RF-DETR distingue mejor la textura/tonalidad de esta lesión, minimizando confusiones con sarna común y *background*.

Sana

- RF-DETR: P = 92,00 %, R = 88,00 %, F1 $\approx$ 94,00 %.
- YOLOv12/YOLOv26: P = 92,00 %, R = 87,00 %, F1 $\approx$ 89,00 %.
- Conclusión: Los tres modelos son robustos; RF-DETR ofrece una ventaja pequeña.

Tabla 5
Mejor modelo por clase

Clase	Mejor <i>precision</i>	Mejor <i>recall</i>	Mejor F1	Comentario clave
Pierna negra	YOLO v26 (78,57 %)	YOLO v12 (75,00 %)	YOLO v26 (~71,00 %)	YOLOv26 balancea; YOLOv12 maximiza sensibilidad.
Podre dumbre rosada	RF-D ETR (100,00 %)	RF-D ETR (75,00 %)	RF-DE TR (~86,00 %)	RF-DETR sin falsos positivos y con alta R.
Pudri ción seca	RF-D ETR (53,00 %)	RF-D ETR (75,00 %)	RF-DE TR (~62,00 %)	RF-DETR maneja bordes difusos/subepidérmicos.
Sarna común	RF-D ETR (50,00 %)	RF-D ETR (100,00 %)	RF-DE TR (~67,00 %)	Reduce sobredetección de YOLO.
Sarna negra	RF-D ETR (93,00 %)	RF-D ETR (82,00 %)	RF-DE TR (~88,00 %)	Distingue mejor textura/tonos oscuros.
Sana	RF-D ETR (92,00 %)	RF-D ETR (88,00 %)	RF-DE TR (~90,00 %)	Diferencias marginales; todos robustos.

Discusión

Panorama general y coherencia con la literatura

Bajo un protocolo común (misma partición, resolución y protocolo de entrenamiento), los tres detectores alcanzaron rendimientos competitivos, con paridad entre YOLOv12 y YOLOv26 en las métricas globales (P/R y mAP) y una ventaja en mAP para RF-DETR (tanto a $\text{IoU} = 0.50$ como en el rango $0.50\text{--}0.95$). Según los registros, YOLOv12 consiguió *precision* = 78,72 %, *recall* = 96,86 %, $\text{mAP}@0.50$ = 79,36 % y $\text{mAP}@0.50:0.95$ = 73,26 %, mientras que YOLOv26 obtuvo *precision* = 79,34 %, *recall* = 96,64 %, $\text{mAP}@0.50$ = 78,48 % y $\text{mAP}@0.50:0.95$ = 73,55 %. En múltiples aplicaciones agrícolas contemporáneas, la familia YOLO se ha impuesto por su equilibrio velocidad-exactitud y adopción práctica en el campo (incluidos tubérculos y tareas en poscosecha), lo que concuerda con revisiones sistemáticas y estudios de *benchmarking* recientes (Badgujar y otros, 2024; Divyanth y otros, 2024).

Por su parte, la familia DETR —y en particular las variantes en tiempo real— prescinde de NMS y aprovecha las consultas (*queries*) con asignación bipartita, lo que a menudo se traduce en mejor consistencia de localización a IoU estrictos y curvas P–R con mayor área (AP) en rangos amplios de umbrales (Dahiya y otros, 2025). Este comportamiento está documentado por RT-DETR (CVPR 2024) (Zhao y otros, 2024) y sus implementaciones, que muestran mejoras conjuntas en velocidad y AP respecto a detectores de una etapa y a otros DETR clásicos. Además,

comienzan a reportarse aplicaciones agrícolas con RF-DETR, reforzando la pertinencia de incluir un *transformer-based* en esta comparación (Blok y otros, 2025; Öztürk & Şin, 2026).

Métricas globales: P/R puntuales vs. AP integrados ($\text{IoU} = 0.50$ y $0.50\text{--}0.95$)

Las métricas puntuales (P y R) dependen del umbral de confianza y del posprocesamiento (p. ej., NMS en YOLO), mientras que $\text{mAP}@0.50$ y $\text{mAP}@0.50:0.95$ integran el área bajo la curva P–R al barrer umbrales (y, en COCO, distintos IoU), ofreciendo una medida umbral-agnóstica de clasificación y localización. Esta diferencia explica por qué un modelo puede exhibir AP superior y, aun así, mostrar P/R inferiores en un punto de operación dado: el óptimo global de la curva puede no coincidir con el umbral usado en validación/inferencia (Thibaut, 2023).

En los resultados, RF-DETR presentó mAP superior tanto a 0.50 como a $0.50\text{--}0.95$, coherente con la literatura que reporta mejor localización a IoU estrictos, eliminación de NMS y mayor área P–R gracias a consultas de mayor calidad y decodificadores eficientes (Parlak, 2026; Rahman Sams y otros, 2026). En cambio, YOLOv12/ YOLOv26 retuvieron R y P puntuales ligeramente mejores al umbral por defecto (Ultralytics), un patrón habitual en detectores de una etapa con NMS y calibración hiperprobada en la práctica agrícola (Badgujar y otros, 2024; Divyanth y otros, 2024).

Desempeño por clase: lectura comparada y causas probables

Pierna negra. YOLOv26 alcanzó el mejor F1 ($P \approx 79,00\%$; $R \approx 65,00\%$), equilibrando precisión y sensibilidad; YOLOv12 retuvo el *recall* más alto ($75,00\%$), idóneo para tamizado. La menor R de RF-DETR sugiere umbral subóptimo o necesidad de muestras adicionales en esta clase concreta. Este tipo de intercambio P–R entre arquitecturas se ha observado en *benchmarks* agrícolas, donde la textura y la humedad influyen en la confusión interclase (Badgujar y otros, 2024).

Podredumbre rosada. RF-DETR fue superior ($P = 100,00\%$; $R = 75,00\%$), esto es, sin falsos positivos y con alta R al punto de operación usado; la curva P–R de esta clase tiende a dominar a lo largo de los umbrales, coherente con las ventajas de los transformadores en contexto y consistencia espacial (Zhao y otros, 2024).

Pudrición seca. RF-DETR superó a YOLO en R y F1; su mejor desempeño ante bordes difusos y manifestaciones subepidérmicas sugiere que la función de coste ($L1+GIoU$) y el matching uno-a-uno amortiguan la penalización a IoU altos; aun así, la literatura coincide en que lesiones internas/incipientes requieren NIR/SW-NIR/HSI para elevar sensibilidad, lo que refuerza una vía multimodal futura (Prasetyo y otros, 2024; F. Zhang y otros, 2024).

Sarna común vs. sarna negra. En YOLO se observa sobredetección en sarna común (P baja, R alta), mientras que RF-DETR mejora el balance (mayores P y R), probablemente por capturar mejor la textura granular y el contexto espacial de los esclerocios de *Rhizoctonia solani* (sarna negra) y su variabilidad (AG-3, etc.), conocidos por inducir ambigüedades visuales en tubérculos (Chaudhary y otros, 2023; *Potato - Rhizoctonia Solani* (Stem Canker & Black Scurf), 2025).

Sana. Los tres modelos son robustos (P y R altos, con ventaja marginal de RF-DETR), resultado coherente con aplicaciones en tubérculos/poscosecha y revisiones sobre YOLO en contextos agrícolas (Badgujar y otros, 2024; Divyanth y otros, 2024).

Validez estadística y transparencia

Dado el tamaño del conjunto de prueba ($n = 57$), las diferencias pequeñas (≤ 1 punto porcentual) pueden no ser estadísticamente significativas. Se recomienda aplicar remuestreo por imagen o instancia para obtener intervalos de confianza del $95,00\%$ en P , R , mAP y F1; así como pruebas de permutación para contrastar la hipótesis nula de “sin diferencia”. Además, se sugiere reportar curvas P–R con puntos de máxima F1 (global y por clase) para anclar el punto de operación a la política de riesgo (costo de FN vs. FP). Estas prácticas son coherentes con la evaluación COCO y con guías recientes para mAP y P–R (Thibaut, 2023; F. Zhang y otros, 2024).

Implicaciones prácticas en TIC

- 1. Arquitectura de sistema (*edge* ↔ nube).

Los resultados sustentan un diseño híbrido: inferencia en el borde (cámaras/SoC/Jetson) para respuesta en tiempo real y envío *event-driven* de casos ambiguos a la nube para auditoría y reetiquetado. La familia YOLO es ampliamente adoptada en el agro por su baja latencia y facilidad de despliegue; RF-DETR aporta detección *end-to-end* sin NMS, lo que simplifica la canalización de posprocesamiento en dispositivos con recursos limitados y ofrece mejor AP a IoU estrictos (Badgujar y otros, 2024; F. Zhang y otros, 2024).

- 2. Selección de modelo por caso de uso (puntos de operación).

Para tamizado de alto volumen (no omitir lesiones), un YOLO calibrado a umbral bajo maximiza *recall*; para diagnóstico con cuantificación de severidad (IoU altos), RF-DETR ofrece mayor mAP@0.50:0.95 y cajas más consistentes a lo largo del barrido de umbrales. Las curvas P-R y mAP (COCO) son la herramienta recomendada para fijar umbrales por clase según costo de error (Thibaut, 2023; Zhao y otros, 2024).

- 3. *Pipelines* en dos etapas (*screening* → refinamiento).

Un flujo YOLO → RF-DETR reduce el costo computacional en *edge* (detección rápida) y eleva la *precision*/IoU en los positivos (refinamiento sin NMS), estrategia ya sugerida en *benchmarking* de visión agrícola y consistente con la evidencia de RT-DETR en tiempo real (Zhou y otros, 2025).

- 4. Multimodalidad para lesiones internas.

Para pudrición seca y otras lesiones subepidérmicas, integrar NIR/SW-NIR/HSI en el *pipeline* TIC eleva la detección temprana, superando las limitaciones de RGB. La fusión tardía (*late fusion*) o el enrutamiento de muestras sospechosas a una estación NIR/HSI son estrategias escalables en plantas con limitaciones de tiempo-ciclo (Prasetyo y otros, 2024; F. Zhang y otros, 2024).

Conclusiones

1. Rendimiento comparativo y viabilidad operativa

Bajo un mismo protocolo experimental — mismo conjunto de datos, partición estratificada y resolución de entrada (640 px)—, los detectores de la familia YOLO y el modelo basado en transformadores RF-DETR demostraron ser viables para la detección temprana de enfermedades en tubérculos de papa en escenarios poscosecha. YOLOv12 y YOLOv26 presentaron un desempeño comparable en términos de *precision* y *recall*, mientras que RF-DETR obtuvo valores superiores de mAP tanto a IoU = 0.50 como en el rango 0.50–0.95. Estos resultados indican la pertinencia de ambos paradigmas para su integración en sistemas automatizados de inspección agrícola con recursos computacionales moderados.

2. Sensibilidad, localización y elección según el caso de uso

El análisis comparativo evidenció un intercambio sistemático entre sensibilidad y exactitud de localización. Los modelos YOLO mostraron un *recall* ligeramente superior al umbral por defecto, lo que los hace adecuados para escenarios de tamizado masivo donde minimizar falsos negativos es prioritario. En contraste, RF-DETR presentó una localización más estricta y consistente, reflejada en su ventaja en mAP@0.50:0.95, atribuible a su mecanismo de asignación uno a uno y a la ausencia de supresión no máxima. Este comportamiento resulta relevante cuando la delimitación precisa de la lesión es necesaria para la estimación de severidad o clasificación posterior.

3. Comportamiento por clase y limitaciones del enfoque RGB

El análisis mediante matrices de confusión permitió identificar patrones de error específicos del dominio agrícola. En particular, las confusiones entre sarna común y sarna negra se explican por similitudes fenotípicas en textura y coloración superficial, reportadas en la literatura fitopatológica. Asimismo, las lesiones subepidérmicas asociadas a *Fusarium spp.* presentaron mayores tasas de falsos negativos y menor solapamiento espacial, lo que evidencia una limitación inherente al uso exclusivo de imágenes RGB para la detección temprana de daños internos.

4. Implicaciones prácticas y recomendación de arquitectura

A partir de los resultados obtenidos, se propone una arquitectura funcional en dos etapas: un primer nivel de detección basado en YOLO, orientado a maximizar la sensibilidad y permitir un cribado rápido, seguido de un segundo nivel con RF-DETR para el refinamiento de la localización y la mejora del IoU. Esta combinación equilibra eficiencia computacional y precisión espacial y puede reforzarse mediante técnicas de calibración y, cuando el objetivo sea reducir omisiones en lesiones incipientes, mediante la fusión tardía con información espectral NIR o HSI. En conjunto, los hallazgos sugieren la viabilidad de desarrollar una herramienta automatizada, precisa y escalable, alineada con los requisitos de sistemas de monitoreo en tiempo real y *edge computing* en entornos poscosecha.

Declaración de conflictos de interés

Quien firma este artículo declara no tener conflictos de interés.

Fuentes de financiación

Quien firma este artículo declara que esta investigación no recibió financiamiento externo

de ninguna agencia de financiamiento del sector público, comercial o sin fines de lucro.

Contribución de autoría (CRediT):

Manolo Sebastián Muñoz-Espinoza: Conceptualización; metodología; software; validación; análisis formal; investigación; recursos; gestión

de datos; redacción - borrador original; redacción - revisión y edición; visualización; supervisión; administración del proyecto.

Referencias

- Aleinzi, M. (2025). Enhancing precision agriculture with YOLOv8: A deep learning approach to potato disease identification. *International Journal of Advanced Computer Science and Applications*, 16(4), 1156–1166.
<https://doi.org/10.14569/IJACSA.2025.01604110>
- Badgujar, C. M., Poulouse, A., & Gan, H. (2024). Agricultural object detection with You Look Only Once (YOLO) algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture*, 223.
<https://doi.org/10.1016/j.compag.2024.109090>
- Blok, P. M., Wang, H., Suh, H. K., Wang, P., Burridge, J., & Guo, W. (2025). PointRAFT: 3D deep learning for high-throughput prediction of potato tuber weight from partial point clouds [Preprint]. *arXiv*.
<https://doi.org/10.48550/arXiv.2512.24193>
- Chaudhary, S., Lal, M., Sagar, S., Sharma, S., & Kumar, M. (2023). Black scurf of potato: Insights into biology, diagnosis, detection, host-pathogen interaction, and management strategies. *Tropical Plant Pathology*, 49(2), 169–192.
<https://doi.org/10.1007/s40858-023-00622-4>
- Dahiya, N., Prakash, D., Kundu, S., Kuttan, S. R., Suwalka, I., Ayadi, M., Dubale, M., & Hashmi, A. (2025). Optimised RFO tuned RF-DETR model for precision urine microscopy for renal and systemic disease diagnosis. *Scientific Reports*, 15(1), 25842.
<https://doi.org/10.1038/s41598-025-11725-0>
- Divyanth, L. G., Khanal, S. R., Paudel, A., Mattupalli, C., & Karkee, M. (2024). Efficient detection of eyes on potato tubers using deep-learning for robotic high-throughput sampling. *Frontiers in Plant Science*, 15, 1512632.
<https://doi.org/10.3389/fpls.2024.1512632>
- Ephytia. (2025, July 24). Potato—Rhizoctonia solani (stem canker & black scurf). <https://ephytia.inra.fr/en/C/20899/Potato-Rhizoctonia-solani-Stem-canker-Black-scurf>
- Faria, F. T. J., Moin, M. Bin, & Wase, A. Al. (2023). *Potato Diseases Datasets* [Conjunto de datos]. Kaggle.
<https://doi.org/10.34740/KAGGLE/DSV/7203264>

- Geoffrey Manurung, D., Ryan Pinasthika, M., Obila Vasya, M. A., Dwi Setya Putri, R. A., Parasian Tampubolon, A., Fadhil Prayata, R., Khoirin Nisa, S., & Yudistira, N. (2024). Deteksi dan klasifikasi hama potato beetle pada tanaman kentang menggunakan YOLOV8. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(4), 723–734.
<https://doi.org/10.25126/jtiik.1148092>
- Guillemette, A. M., Casanova, G. H., Hamilton, J. P., Pokorná, E., Dobrev, P. I., Motyka, V., Rashotte, A. M., & Leisner, C. P. (2024). The physiological and molecular responses of potato tuberization to projected future elevated temperatures. *Plant Physiology*, 197(1), 664.
<https://doi.org/10.1093/PLPHYS/KIAE664>
- He, Y., Peng, Y., Wei, C., Zheng, Y., Yang, C., & Zou, T. (2024). Automatic disease detection from strawberry leaf based on improved YOLOv8. *Plants*, 13(18), 2256
<https://doi.org/10.3390/PLANTS13182556>
- Jayanthi, J., & Kumar, K. A. (2024). Transformative impact of AI-driven computer vision in agriculture. En *Artificial intelligence techniques in smart agriculture* (pp. 129–150). Springer
https://doi.org/10.1007/978-981-97-5878-4_9
- Jocher, G., & Qiu, J. (2026). *Ultralytics YOLO26* [Software]. GitHub.
<https://github.com/ultralytics/ultralytics>
- Keithellakpam, L. B., Karunakaran, C., Singh, C. B., Jayas, D. S., & Danielski, R. (2026). A comprehensive review on pre- and post-harvest perspectives of potato quality and non-destructive assessment approaches. *Applied Sciences*, 16(1), 190
<https://doi.org/10.3390/APP16010190>
- Liu, C. (2024). Advancing strawberry disease detection in agriculture: A transfer learning approach with YOLOv5 algorithm. *International Journal of Advanced Computer Science and Applications*, 15(3), 1013–1022.
<https://dx.doi.org/10.14569/IJACSA.2024.01503101>
- Ma, P., Li, C., Rahaman, M. M., Yao, Y., Zhang, J., Zou, S., Zhao, X., & Grzegorzec, M. (2022). A state-of-the-art survey of object detection techniques in microorganism image analysis: From classical methods to deep learning approaches. *Artificial Intelligence Review*, 56(2), 1627–1698.
<https://doi.org/10.1007/S10462-022-10209-1>

- Muñoz Espinoza, M., Moreno Castillo, W. E., y Palacios Ruiz, F. P. (2026). Detección de enfermedades en papa mediante deep learning usando YOLOv12. *Ciencias de la Ingeniería y Aplicadas*, 10(1), 119–129.
<https://doi.org/10.61236/CIYA.V10I1.1239>
- Moreno Mayhuire, J. S., & Herrera Quispe, J. A. (2023). Modelos de aprendizaje automático para la clasificación de enfermedades de la *Solanum tuberosum*: Una revisión sistemática de la literatura. *Revista de Investigación de Sistemas e Informática*, 16(2), 129–137.
<https://doi.org/10.15381/risi.v16i2.25980>
- Nagel, M., Dulloo, M. E., Bissessur, P., Gavrilenko, T., Bamberg, J., Ellis, D., & Giovannini, P. (2022). *Global strategy for the conservation of potato*. Leibniz Institute of Plant Genetics and Crop Plant Research.
<https://doi.org/10.5447/IPK/2022/29>
- Öztürk, L., & Şin, B. (2026). Real-time transformer-based object detection for advanced plant disease recognition: The grapevine fanleaf virus example. *Applied Fruit Science*, 68(1).
<https://doi.org/10.1007/s10341-026-01777-5>
- Palacio Betancur, S., & Bolaños Martínez, F. (2025). Tomato (*Solanum lycopersicum* L.) leaf disease detection using computer vision. *Revista Facultad Nacional de Agronomía*, 78(3), 11203–11212.
<https://doi.org/10.15446/rfnam.v78n3.116493>
- Parlak, C. (2026). Evaluation of cutting-edge object detection architectures on multi-object and single-object datasets. *Black Sea Journal of Engineering and Science*, (9)1. 287-294.
<https://doi.org/10.34248/bsengineering.1736319>
- Prasetyo, E. W., Amanah, H. Z., Farras, I., Pahlawan, M. F. R., & Masithoh, R. E. (2024). Detection of *Fusarium spp.* infection in potato (*Solanum tuberosum* L.) during postharvest storage through visible-near-infrared and shortwave-near-infrared reflectance spectroscopy. *Open Agriculture*, 9(1).
<https://doi.org/10.1515/opag-2022-0295>
- Qin, R., Wang, Y., Liu, X., & Yu, H. (2024). Advancing precision agriculture with deep learning enhanced SIS-YOLOv8 for Solanaceae crop monitoring. *Frontiers in Plant Science*, 15, 1485903.
<https://doi.org/10.3389/fpls.2024.1485903>
- Rahaman Sams, F., Kazi Supti, S., Binte Hamid, S., Junayed, R., Fahim A Bari, K. M., Junaeid Ali, M., Gani, R., Shams, K., Rifat Ahmmad Rashid, M., & Ul Islam, R. (2026). Real-time sunflower detection using semi-supervised and self-supervised deep learning for precision agriculture. *Smart Agricultural Technology*, 13, 101684.
<https://doi.org/10.1016/J.ATECH.2025.101684>

- Redmon, J. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Robinson, I., Robicheaux, P., Popov, M., Ramanan, D., & Peri, N. (2026). RF-DETR: Neural architecture search for real-time detection transformers [Preprint]. arXiv.
<https://arxiv.org/abs/2511.09554>
- Roboflow. (2025). *Roboflow: Computer vision tools for developers and enterprises*.
<https://roboflow.com/>
- Sapkota, R., Cheppally, R. H., Sharda, A., & Karkee, M. (2025). YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection [Preprint]. arXiv.
<https://arxiv.org/pdf/2509.25164>
- Sharma, S., Tiwari, R. K., Sagar, V., & Maharana, C. (2024). Soil- and tuber-borne diseases of potato. In *Approaches for potato crop improvement and stress management* (pp. 179–231).
https://doi.org/10.1007/978-981-97-1223-6_7
- Solankey, S. S. (Ed.). (2025). *Advances in research on potato production*. Springer.
<https://doi.org/10.1007/978-3-031-82710-5>
- Thibaut, L. (2023, August 13). *COCO evaluation metrics explained*. Picsellia.
<https://www.picsellia.com/post/coco-evaluation-metrics-explained>
- Tian, Y., Ye, Q., & Doermann, D. (2025, February 20). YOLO12: Attention-centric real-time object detectors [Preprint]. arXiv.
<https://arxiv.org/abs/2502.12524>
- Upeniece, L., Bimšteine, G., & Stramkale, V. (2025). The most important potato diseases and microbiological agents for their control: A review. *Research for Rural Development 2025: Annual 31st International Scientific Conference Proceedings* (Vol. 1, pp. 17–23). Latvia University of Life Sciences and Technologies.
<https://doi.org/10.22616/RRD.31.2025.003>
- Yurdakul, M., Sazak, H., Kotan, M., & Taşdemir, T. (2026). A review of YOLO family from YOLOv1 to YOLO26 [Preprint]. Preprints.org
<https://doi.org/10.20944/PREPRINTS202602.1844.V1>

- Zhang, F., Wang, W. X., Wang, C. S., Zhou, J., Pan, Y., & Sun, J. F. (2024). Study on hyperspectral detection of potato dry rot in gley stage based on convolutional neural network. *Spectroscopy and Spectral Analysis*, 44(2), 480-489.
[https://doi.org/10.3964/j.issn.1000-0593\(2024\)02-0480-10](https://doi.org/10.3964/j.issn.1000-0593(2024)02-0480-10)
- Zhang, J., Xie, J., Zhang, F., Gao, J., Yang, C., Song, C., Rao, W., & Zhang, Y. (2024). Greenhouse tomato detection and pose classification algorithm based on improved YOLOv5. *Computers and Electronics in Agriculture*, 216, 108519.
<https://doi.org/10.1016/J.COMPAG.2023.108519>
- Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., & Chen, J. (2024). DETRs beat YOLOs on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16965–16974).
<https://zhao-yian.github.io/RTDETR>.
- Zhou, H., Liu, Y., & Wang, J. (2025). Railway fastener defect detection using RFD-DETR: A lightweight real-time transformer-based approach. *PLOS ONE*, 20(11), e0331513.
<https://doi.org/10.1371/JOURNAL.PONE.0331513>