

Detección de enfermedades en papa mediante deep learning usando YOLOv12

The Disease detection in potatoes using deep learning with YOLOv12

Manolo Muñoz-Espinoza¹, Walter Eduardo-Moreno Castillo², Franck Pío Palacios-Ruiz³

¹ Universidad de Holguí. Email: arabel@facing.uho.edu.cu

² Universidad de Holguín. Email: robertos@facing.uho.edu.cu

³ Universidad de Holguín. Email: mabele@facing.uho.edu.cu

DOI: <https://doi.org/10.61236/ciya.v10i1.1239>

Recibido: 25 de octubre de 2025 • Revisión aceptada: 5 de enero de 2026

Resumen

La papa es un cultivo alimenticio esencial, pero su producción se ve amenazada por enfermedades como tizón temprano y tardío, cuya detección tardía genera pérdidas económicas y ambientales significativas. La detección temprana y precisa de enfermedades en cultivos es fundamental para garantizar la seguridad alimentaria y la sostenibilidad agrícola. En este trabajo, se propone un enfoque innovador basado en deep learning para la identificación automática de enfermedades en hojas de papa (*Solanum tuberosum*), mediante una arquitectura mejorada de YOLOv12 que sustituye parcialmente las redes neuronales convolucionales tradicionales por mecanismos de atención. El modelo fue entrenado con recursos computacionales accesibles (2× GPU NVIDIA T4 en Kaggle) y un conjunto de datos inicial compuesto por 363 imágenes originales (121 por clase: tizón temprano, tizón tardío y hojas sanas), ampliado a 920 mediante un pipeline de aumento de datos realista (rotación, flip, shear, ruido y variación de exposición). Gracias a la implementación de early stopping, el entrenamiento converge eficientemente en 43 épocas, alcanzando un rendimiento sobresaliente: precisión = 0,9854, mAP50 = 0,9950 y recall = 1,0000, estos indicadores demuestran alta sensibilidad y especificidad, incluso con datos limitados. En conclusión, el sistema propuesto es una solución viable, robusta y escalable para su integración en aplicaciones de agricultura de precisión, facilitando diagnósticos tempranos y reduciendo la dependencia de inspecciones manuales.

Palabras Clave: Papa, enfermedades, detección, deep learning, YOLO, agricultura de precisión.

Abstract

Potato is an essential food crop, but its production is threatened by diseases such as early and late blight, whose delayed detection leads to significant economic and environmental losses. Early and accurate disease detection in crops is crucial for ensuring food security and agricultural sustainability. In this work, we propose an innovative deep learning-based approach for the automatic identification of diseases in potato leaves (*Solanum tuberosum*) using the standard YOLOv12 architecture. The model was trained on accessible computational resources (2× NVIDIA T4 GPUs on Kaggle) using an initial dataset of 363 original images (121 per class: early blight, late blight, and healthy leaves), expanded to 920 images through a realistic data augmentation pipeline (including rotation, flipping, shear, noise, and exposure variation). Thanks to early stopping, training converged efficiently within 43 epochs, achieving outstanding performance: precision = 0,9854, mAP50 = 0,9950, and recall = 1,0000. These metrics demonstrate high sensitivity and specificity, even with limited data. In conclusion, the proposed system represents a viable, robust, and scalable solution for integration into precision agriculture applications, enabling timely diagnoses and reducing reliance on manual inspections.

Key words: Potato, diseases, detection, deep learning, YOLO, precision agriculture.

I. INTRODUCCIÓN

La agricultura constituye uno de los pilares fundamentales de la economía global y desempeña un papel crítico en la seguridad alimentaria y la sostenibilidad de los sistemas agroalimentarios [1]. En este contexto, la papa (*Solanum tuberosum* L.) se posiciona como el cuarto cultivo alimentario más importante del mundo en términos de producción, después del maíz, trigo y arroz [2], y representa una fuente esencial de nutrientes y calorías para más de mil millones de personas, especialmente en regiones en desarrollo [3] [4]. No obstante, este cultivo enfrenta amenazas constantes por enfermedades fitosanitarias de origen fúngico, las cuales pueden ocasionar pérdidas catastróficas en rendimiento y calidad, comprometiendo la estabilidad de los medios de vida rurales y la disponibilidad de alimentos [5].

Tradicionalmente, la identificación de enfermedades foliares ha dependido de inspecciones visuales por expertos, un enfoque que resulta lento, subjetivo, costoso y poco escalable [6] [7]. En respuesta a estas limitaciones, la agricultura de precisión ha emergido como un paradigma transformador, integrando tecnologías digitales para monitorear, analizar y gestionar los cultivos de manera eficiente y sostenible [8] [9]. En los últimos años, el aprendizaje profundo (deep learning) y el análisis de grandes volúmenes de datos (big data) han demostrado un potencial

extraordinario en aplicaciones agrícolas, particularmente en la detección automatizada de enfermedades en plantas a partir de imágenes foliares [10] [11]. Estos enfoques permiten no solo identificar patologías con alta precisión, sino también localizarlas espacialmente en tiempo real, facilitando intervenciones oportunas y localizadas.

En este contexto, los modelos de detección de objetos en tiempo real, en particular la arquitectura YOLO (You Only Look Once) [12], han ganado relevancia por su equilibrio entre precisión, velocidad y eficiencia computacional, cualidades esenciales para su implementación en entornos agrícolas reales [13]. Sin embargo, su efectividad se ve frecuentemente limitada por la escasez de datos etiquetados, un problema crítico en cultivos específicos como la papa, donde la adquisición y anotación de imágenes requiere expertise fitopatológico y condiciones controladas [14].

Este trabajo aborda dicha limitación mediante el desarrollo de un pipeline de aumento de datos estratégico y realista, diseñado específicamente para simular las variaciones naturales del entorno de campo (rotación, iluminación, ruido, entre otros), a partir de un conjunto inicial reducido de sólo 121 imágenes por clase. El objetivo principal de este estudio es demostrar que, mediante un diseño cuidadoso del preprocesamiento y la selección del modelo, es posible entrenar un sistema robusto basado en YOLOv12 para la detección y localización precisa del tizón

temprano, el tizón tardío y hojas sanas en papa, incluso con datos escasos. Los resultados presentados sientan las bases para soluciones escalables, accesibles y de bajo costo en el monitoreo fitosanitario de cultivos estratégicos.

II. METODOLOGÍA

En esta sección se describe la metodología empleada para el desarrollo y evaluación del modelo propuesto, el proceso comprende adquisición de los datos, etiquetado, preprocesamiento, aumento de datos, división del dataset, entrenamiento y resultados, como se indica en la tabla 1.

Adquisición de Datos	Etiquetado	Procesamiento - Aumento
		Resize: 640 x 640 Orientación automática
		Flip, Rotation Shear, Noise Bounding Box: Flip
		Bounding Box: Exposure

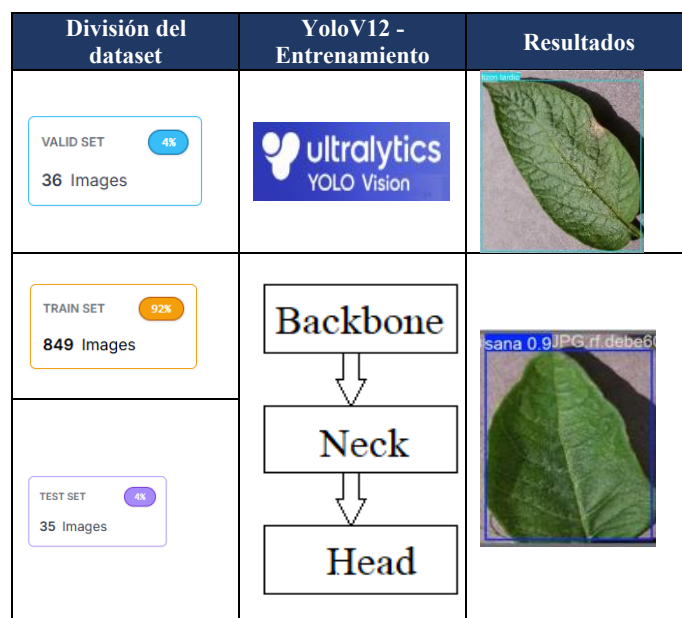


Tabla 1. Propuesta metodológica para detección de enfermedades multiclase en papa con YOLOv12

2.1 Adquisición de datos

Las fotografías se obtuvieron de Kaggle, el dataset es llamado PlantVillage [15], en este caso se obtuvo solo las fotografías del cultivo de papa, con dos enfermedades y hojas sanas, se escogieron el mismo número de imágenes para que el entrenamiento sea más efectivo (tabla 2).

Tabla 2. Distribución de imágenes y clase

Clase	Número de imágenes
Sana:	121
Tizón tardío	121
Tizón temprano	121

2.2 Etiquetado de imágenes

Mediante la plataforma de Roboflow [16], se realizó el etiquetado a mano de cada imagen con su respectiva clase, esto consistió en dibujar el bounding box indicando en que sitio de la fotografía se encontraba la hoja.

2.3 Procesamiento y aumento de datos

Una vez que todas las imágenes fueron etiquetadas, se procedió al redimensionamiento a 624x624 píxeles de la imagen original, debido a que la arquitectura YOLO necesita esta resolución para las imágenes de entrada.

Luego en la misma plataforma Roboflow, se hizo el aumento de datos:

- **Flip:** Horizontal, Vertical
- **Rotation:** Between -15° and $+15^\circ$
- **Shear:** $\pm 24^\circ$ Horizontal, $\pm 34^\circ$ Vertical
- **Noise:** Up to 1,92% of pixels
- **Bounding Box: Flip:** Horizontal
- **Bounding Box: Exposure:** Between -10% and +10%

2.4 División del dataset

Con el aumento de datos, el dataset final se muestra en la Tabla 3:

Tabla 3. Dataset final

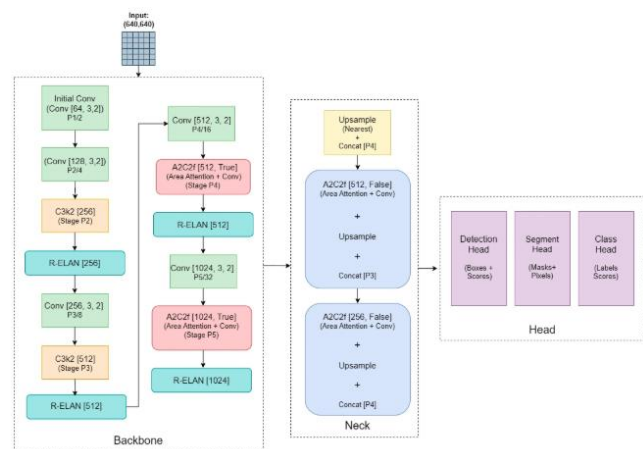
División	Número de imágenes / Porcentaje
Train	849 / 92%
Test	36 / 4%
Valid	35 4%

2.5 Modelo YOLOv12

La arquitectura YOLOv12 es la última versión de Ultralytics que resuelve tareas sobre detección, segmentación, clasificación, pose y la detección de objetos orientados OBB. La arquitectura YOLOv12 introduce un cambio paradigmático respecto a versiones anteriores al reemplazar los bloques convolucionales tradicionales por

mecanismos de atención. Esta estrategia permite al modelo priorizar las regiones más relevantes de la imagen, optimizando así la extracción de características y aumentando la precisión en la detección, todo ello manteniendo la eficiencia necesaria para el procesamiento en tiempo real [17].

Tabla 4. Arquitectura YOLOv12 [18]



2.6 Entrenamiento del modelo

El entrenamiento del modelo se realizó utilizando la infraestructura proporcionada por la plataforma Kaggle, empleando dos unidades de procesamiento gráfico (GPU) del tipo NVIDIA T4, esta configuración permite acelerar significativamente los cálculos necesarios para el entrenamiento del modelo de deep learning, reduciendo el tiempo total de cómputo y facilitando la convergencia eficiente del algoritmo.

Los parámetros de entrenamiento se muestran en la tabla 4.

Tabla 5. Parámetros de entrenamiento YOLOv12

Parámetros	Valor de entrada
Dataset inicial	363 imágenes
Data augmentation	920 imágenes
Tamaño del modelo	YOLOv12 Nano – Turbo – 2.5 M parámetros
Epochs	150
Optimizador	AdamW
Batch Size	16
Número de clases	3
Learning Rate	0,00140143

El conjunto de datos original proviene de Kaggle y contiene 363 imágenes de plantas de papa con distintas condiciones. Dado que este número es relativamente pequeño para entrenar un modelo de inteligencia artificial, se aplicó la aumentación de datos explicado en 2.3, llegando a un total de 920 imágenes.

Se eligió el modelo YOLOv12 Nano – Turbo, una versión ligera y rápida de YOLO, ideal para aplicaciones prácticas en el campo, ya que funciona bien incluso en dispositivos con pocos recursos, como celulares o computadoras básicas. Este modelo tiene solo 2.5 millones de parámetros, lo que lo hace eficiente sin perder demasiada precisión.

El entrenamiento se realizó durante 150 epochs (vueltas completas sobre el conjunto de datos), lo suficiente para que el modelo aprendiera bien sin memorizar los ejemplos. Se usó el optimizador AdamW, que ayuda a que el entrenamiento sea estable y eficaz, y un tamaño de lote (batch size) de 16, adecuado para el equipo utilizado.

El problema se planteó con 3 clases: tizón tardío, mancha temprana y papa sana, ya que son las categorías más relevantes para los agricultores en la práctica.

La tasa de aprendizaje se ajustó a 0.00140143 tras pruebas preliminares, buscando un equilibrio entre rapidez de aprendizaje y estabilidad durante el entrenamiento.

III. ANÁLISIS DE RESULTADOS

El modelo YOLOv12 fue entrenado durante 43 épocas, deteniéndose automáticamente mediante early stopping al no observarse mejoras significativas en las métricas de validación. Este comportamiento refleja la eficiencia del modelo y su capacidad para converger rápidamente sin sobreajuste, lo cual es especialmente relevante en aplicaciones agrícolas donde los recursos computacionales y los conjuntos de datos pueden ser limitados.

3.1 Matriz de confusión: Evaluación por clase

La matriz de confusión (Figura 1), proporciona una visión detallada del desempeño del modelo a nivel de clase, permitiendo identificar posibles errores de clasificación entre enfermedades similares. Como se observa, el modelo exhibe un alto grado de precisión en todas las categorías, con muy pocos casos mal clasificados. Esto sugiere que los mecanismos de atención implementados en la arquitectura logran capturar características discriminantes específicas de cada enfermedad, incluso en presencia de síntomas visuales superpuestos o variaciones en la iluminación y orientación de las imágenes.

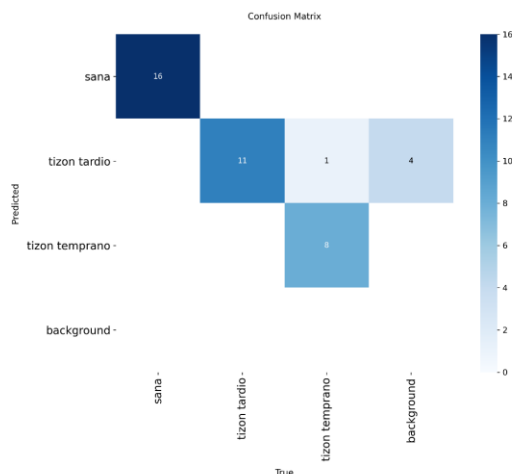


Figura 1. Matriz de confusión.

Resultados por clase:

Clase “sana”

Predicciones correctas: 16

Total, real de “sana”: 16

- Exactitud (precisión) = $16/16 = 100\%$
- Recall (sensibilidad) = $16/16 = 100\%$

El modelo identificó perfectamente todas las imágenes de plantas sanas. Esto es muy importante, ya que evita falsos alarmas en cultivos saludables.

Clase “tizón tardío”

Predicciones correctas: 11

Erróneas: 1 (clasificado como tizón temprano) + 4 (clasificado como fondo) = 5 errores

Total, real de “tizón tardío”: $11 + 1 + 4 = 16$ casos

- Exactitud (precisión) = $11 / (11 + 0 + 0) = 100\%$ (solo contamos predicciones de esta clase, y todas fueron correctas)
- Recall (sensibilidad) = $11 / 16 = 68.75\%$

El modelo reconoció bien los casos de tizón tardío cuando los detectó, pero falló en 5 de 16 casos (31.25%). Esto sugiere que el modelo aún tiene dificultades para detectar algunos síntomas sutiles o parcialmente visibles de esta enfermedad.

Clase “tizón temprano”

Predicciones correctas: 8

Erróneas: 1 (clasificado como tizón tardío)

Total real de “tizón temprano”: $8 + 1 = 9$ casos

- Exactitud (precisión) = $8 / (8 + 0) = 100\%$
- Recall (sensibilidad) = $8 / 9 = 88.89\%$

El modelo tuvo un buen desempeño en esta clase, con solo 1 error (un caso confundido con tizón tardío). Esto indica que el modelo puede distinguir bien entre ambas enfermedades, aunque no perfectamente.

Exactitud global (accuracy) = $(16 + 11 + 8) / 40 = 35 / 40 = 87.5\%$

3.2 Tabla de métricas clave: Rendimiento global del modelo

En la Tabla 6 se resumen las métricas de rendimiento más relevantes obtenidas durante la fase de validación:

Tabla 6. Rendimiento global del modelo.

Época	precision	recall	mAP50
43	0,98544	1	0,995

Estos valores confirman que el modelo no solo detecta con alta precisión las regiones afectadas (precisión), sino que también es capaz de localizar prácticamente todas las instancias reales (recall = 1.0).

Además, el alto valor de mAP50:0.995, sugiere que el modelo mantiene su precisión incluso bajo umbrales de intersección más exigentes, lo que lo posiciona como una herramienta viable para aplicaciones de campo donde se requiere alta fiabilidad.

3.3 Ejemplos visuales de detección: Validación cualitativa

Para complementar el análisis cuantitativo, se presentan en la figura 2 ejemplos representativos de inferencias realizadas por el modelo sobre imágenes de prueba. En ellos se observa cómo el modelo localiza con precisión las áreas afectadas por enfermedades, asignando etiquetas correctas y generando bounding boxes ajustados a los bordes de los síntomas.

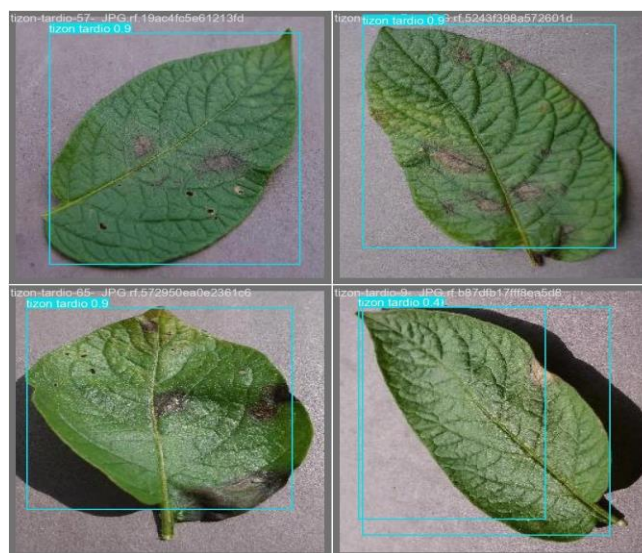


Figura 2. Resultados de predicciones.

Estos resultados cualitativos confirman la capacidad del modelo para generalizar a nuevas muestras, incluso en condiciones de iluminación variable, ángulos de toma distintos o presencia de ruido visual. Además, demuestran que el uso de mecanismos de atención permite al modelo enfocarse en regiones críticas de la imagen, mejorando la extracción de características relevantes para la detección.

IV. CONCLUSIONES

En este estudio se desarrolló e implementó un modelo de detección de enfermedades en papa basado en una arquitectura avanzada de YOLOv12, centrada en mecanismos de atención en lugar de depender exclusivamente de capas convolucionales. Los resultados obtenidos demuestran que este enfoque no solo mantiene, sino que mejora significativamente el rendimiento en comparación con modelos tradicionales, logrando una precisión excepcional = 0.9854,

mAP50 = 0.9950, una sensibilidad perfecta (recall = 1.0) y un accuracy de 87.5%, lo que garantiza que ninguna instancia de enfermedad pase desapercibida.

El uso de early stopping permitió optimizar el proceso de entrenamiento, reduciendo el tiempo de cómputo y evitando el sobreajuste, mientras que la infraestructura de Kaggle con GPU T4 demostró ser suficiente para lograr resultados de alto nivel sin requerir hardware especializado costoso. Esto refuerza la viabilidad de soluciones de deep learning accesibles para pequeños y medianos agricultores.

Además, la evaluación cualitativa mediante ejemplos visuales y la matriz de confusión confirma la capacidad del modelo para generalizar y detectar correctamente múltiples tipos de enfermedades en condiciones diversas, 16/16 plantas sanas bien detectadas, 8/9 casos de tizón temprano identificados y 11/16 de tizón tardío reconocidos.

Estas características lo posicionan como una solución práctica para su integración en aplicaciones móviles, drones o sistemas de monitoreo en campo.

V. BIBLIOGRAFÍA

[1] L. El Hoummaidi, A. Larabi, and K. Alam, "Using unmanned aerial systems and deep learning for agriculture mapping in Dubai,"

Heliyon, vol. 7, no. 10, p. e08154, Oct. 2021, doi: 10.1016/j.heliyon.2021.e08154.

[2] M. E. Calskan, A. Bakhsh, and K. Jabran, Potato production worldwide. Academic Press, 2022.

[3] A. Devaux et al., "Global food security, contributions from sustainable potato agri-food systems," The Potato Crop: Its Agricultural, Nutritional and Social Contribution to Humankind, pp. 3–35, Jan. 2019, doi: 10.1007/978-3-030-28683-5_1/FIGURES/8.

[4] T. Adekanmbi, X. Wang, S. Basheer, S. Liu, A. Yang, and H. Cheng, "Climate change impacts on global potato yields: a review," Environmental Research: Climate, vol. 3, no. 1, p. 012001, Dec. 2023, doi: 10.1088/2752-5295/AD0E13.

[5] G. Alizadeh-Moghaddam, M. Nasr-Esfahani, A. Nasr-Esfahani, E. Sedaghatfar, H. Rahanandeh, and M. J. Yazdi, "Comparative genetic defence analysis using microsatellite markers and anatomical resistance of potato cultivars to early blight," Physiol Mol Plant Pathol, vol. 133, p. 102374, Sep. 2024, doi: 10.1016/J.PMPP.2024.102374.

[6] J. Singh et al., "Transition from conventional to AI-based methods for detection of foliar disease symptoms in vegetable crops: a comprehensive review," Journal of Plant Pathology, pp. 1–24, Aug. 2025, doi: 10.1007/S42161-025-01983-2/METRICS.

- [7] M. Dang et al., “Computer Vision for Plant Disease Recognition: A Comprehensive Review,” *Botanical Review*, vol. 90, no. 3, pp. 251–311, Sep. 2024, doi: 10.1007/S12229-024-09299-Z/METRICS.
- [8] N. Khan and A. Babar, “Innovations in precision agriculture and smart farming: Emerging technologies driving agricultural transformation,” <https://doi.org/10.1142/S2737599424300046>, vol. 11, Jan. 2025, doi: 10.1142/S2737599424300046.
- [9] M. Woźniak and M. F. Ijaz, “Editorial: Recent advances in big data, machine, and deep learning for precision agriculture,” 2024, *Frontiers Media SA*. doi: 10.3389/fpls.2024.1367538.
- [10] R. Sharma, “Artificial intelligence in agriculture: A review,” *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, pp. 937–942, May 2021, doi: 10.1109/ICICCS51141.2021.9432187.
- [11] Q. Zhou, H. Zhang, and S. Wang, “Artificial intelligence, big data, and blockchain in food safety,” *International Journal of Food Engineering*, vol. 18, no. 1, pp. 1–14, Jan. 2022, doi: 10.1515/IJFE-2021-0299/MACHINEREADABLECITATION/RIS.
- [12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016, Accessed: Sep. 24, 2025. [Online]. Available: <http://pjreddie.com/yolo/>
- [13] T. Li, L. Zhang, and J. Lin, “Precision agriculture with YOLO-Leaf: advanced methods for detecting apple leaf diseases,” *Front Plant Sci*, vol. 15, p. 1452502, Oct. 2024, doi: 10.3389/FPLS.2024.1452502/BIBTEX.
- [14] A. Taha and W. Badawy, “Integrating Deep AI with Plant Disease Diagnosis: Toward Early Detection and Sustainable Crop Protection,” *Journal of Scientific Research in Science*, vol. 42, no. 0, pp. 29–47, Aug. 2025, doi: 10.21608/JSRS.2025.393968.1180.
- [15] A. Ali, “PlantVillage Dataset.” Accessed: Sep. 24, 2025. [Online]. Available: <https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset>
- [16] “Roboflow: Computer vision tools for developers and enterprises.” Accessed: Jan. 12, 2025. [Online]. Available: <https://roboflow.com/>
- [17] Y. Tian, Q. Ye, and D. Doermann, “YOLOv12: Attention-Centric Real-Time Object Detectors.” Accessed: Sep. 28, 2025. [Online]. Available: <https://docs.ultralytics.com/es/models/yolo12>
- [18] G. Al-Kateb, M. M. Mijwil, M. Aljanabi, M. Abotaleb, S. R. K. Priya, and P. Mishra, “AI-

PotatoGuard: Leveraging Generative Models for Early Detection of Potato Diseases,” *Potato Res*, vol. 68, no. 1, pp. 449–463, Mar. 2025, doi: 10.1007/S11540-024-09751-Y/METRICS.

[19] A. Abbas, U. Maqsood, S. Ur Rehman, K. Mahmood, T. Alsaedi, and M. Kundi, “An Artificial Intelligence Framework for Disease Detection in Potato Plants,” *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12628–12635, Feb. 2024, doi: 10.48084/ETASR.6456.

[20] H. Afzaal et al., “Detection of a Potato Disease (Early Blight) Using Artificial Intelligence,” *Remote Sensing* 2021, Vol. 13, Page 411, vol. 13, no. 3, p. 411, Jan. 2021, doi: 10.3390/RS13030411.

[21] J. Li, D. Ribeiro, D. Tavares, E. Tiradentes, F. Santos, and D. Rodriguez, “Performance Evaluation of YOLOv11 and YOLOv12 Deep Learning Architectures for Automated Detection and Classification of Immature Macauba (*Acrocomia aculeata*) Fruits,” *Agriculture* 2025, Vol. 15, Page 1571, vol. 15, no. 15, p. 1571, Jul. 2025, doi: 10.3390/AGRICULTURE15151571.