



Artículo de revisión / Review article

Visión por computadora para detectar enfermedades fúngicas en cacao: revisión sistemática de arquitecturas de deep learning

Computer vision for detecting fungal diseases in cocoa: a systematic review of deep learning architectures

Yomer Ysidro-Santos ^{1*}; Andy Roel Vasquez-Bueno ¹; Isait Camizán-Ojeda ¹

¹ Universidad Peruana Unión, Tarapoto, Perú

Recibido: 19/04/2026

Aceptado: 01/06/2026

Publicado: 25/07/2026

*Autor de correspondencia: yomer.ysidro@upeu.edu.pe

Resumen: Esta revisión sistemática de la literatura analiza 43 estudios científicos sobre la aplicación de técnicas de aprendizaje profundo (DL) para la detección temprana de enfermedades del cacao (*Theobroma cacao*). Mediante la metodología PRISMA y el marco PICOC, se examinaron modelos y arquitecturas aplicados a tareas de clasificación, detección y segmentación. Los resultados muestran que los modelos ligeros de la familia YOLO alcanzan precisiones de hasta el 96 % en la identificación de enfermedades del cacao, con capacidad para realizar inferencias en tiempo real. Asimismo, las arquitecturas Vision Transformer (ViT) y U-Net registran precisiones entre el 97 % y 99 % en entornos controlados y conjuntos de datos agrícolas relacionados, evidenciando su potencial para aplicaciones basadas en transferencia de aprendizaje. No obstante, persisten limitaciones relacionadas con la generalización entre dominios, la explicabilidad de los modelos y la disponibilidad de datos obtenidos en condiciones reales de campo. Se concluye que futuras investigaciones deben ampliar los conjuntos de datos públicos, incorporar inteligencia artificial explicable (XAI) y aprendizaje autosupervisado (SSL), así como optimizar modelos para dispositivos con recursos limitados, favoreciendo soluciones sostenibles y escalables para el cultivo inteligente del cacao.

Palabras clave: detección de enfermedades del cacao; aprendizaje profundo; yolo; vision transformer; visión por computadora

Abstract: This systematic literature review analyzes 43 scientific studies on the application of deep learning (DL) techniques for the early detection of diseases in cocoa (*Theobroma cacao*). Using the PRISMA methodology and the PICOC framework, models and architectures applied to classification, detection, and segmentation tasks were examined. The results show that lightweight models from the YOLO family achieve accuracy rates of up to 96% in cocoa disease identification, with the capability to perform real-time inference. Likewise, Vision Transformer (ViT) and U-Net architectures achieve accuracy rates ranging from 97% to 99% in controlled environments and related agricultural datasets, demonstrating their potential for transfer learning applications. However, limitations remain regarding cross-domain generalization, model explainability, and the availability of data collected under real-world field conditions. It is concluded that future research should expand public datasets, incorporate explainable artificial intelligence (XAI) and self-supervised learning (SSL), and optimize models for resource-constrained devices, thereby promoting sustainable and scalable solutions for smart cocoa cultivation.

Keywords: cocoa disease detection; deep learning; yolo; vision transformer; computer vision

1. Introducción

Las enfermedades fúngicas del cacao ocasionan pérdidas importantes en la producción mundial cada año, lo que representa un serio riesgo económico y social para millones de pequeños agricultores en regiones tropicales (Kumi et al., 2022; Borromeo et al., 2025). Patologías como la moniliasis (*Moniliophthora roreri*), la escoba de bruja (*Moniliophthora perniciosa*) y la pudrición negra (*Phytophthora* spp.) afectan tanto el rendimiento como la calidad del grano, reduciendo la competitividad del sector y poniendo en riesgo la sostenibilidad de la cadena de valor del cacao. Los métodos de diagnóstico tradicionales, basados principalmente en la inspección visual (Arakeri et al., 2024), presentan limitaciones debido a su carácter subjetivo, la dependencia de la experiencia del productor y su baja escalabilidad en contextos de producción masiva (Jackulin & Murugavalli, 2022). Esta situación hace necesaria la incorporación de herramientas tecnológicas que permitan una detección más temprana, objetiva y eficiente.

La evolución de las técnicas de visión por computadora y Deep Learning (DL) ha impulsado soluciones más precisas y escalables para el diagnóstico agrícola. Inicialmente, las redes neuronales convolucionales (CNN) se aplicaron con éxito en tareas de clasificación de síntomas; posteriormente, detectores como YOLO incorporaron capacidades de localización en tiempo real, mientras que modelos de segmentación como U-Net permitieron delimitar las áreas afectadas. En los últimos años, los Vision Transformers (ViTs) han adquirido relevancia al incorporar mecanismos de atención global (Ruhad et al., 2025; Zahra et al., 2024). En el caso del cacao, se han desarrollado experiencias como Cocoa Companion, una aplicación móvil basada en CNN para el diagnóstico de enfermedades en hojas y frutos (Kumi et al., 2022). Asimismo, se han empleado CNN para clasificar los estados de madurez de las mazorcas de cacao (Borromeo et al., 2025). En otros cultivos, los algoritmos basados en YOLO han mostrado un desempeño favorable para la detección de enfermedades (Paul et al., 2025), mientras que arquitecturas basadas en ViTs e híbridos han demostrado un elevado potencial para la clasificación de enfermedades vegetales (Bhowmik et al., 2024; Vallabhajosyula et al., 2024).

En este contexto, la presente revisión sistemática busca responder a las siguientes preguntas de investigación: (i) ¿qué arquitecturas de Deep Learning se han aplicado en la detección de enfermedades en cacao y cuáles han mostrado mejores resultados?; (ii) ¿cómo se han abordado los desafíos asociados con la variabilidad de las imágenes agrícolas, incluyendo condiciones de iluminación y oclusiones?; y (iii) ¿qué aportes ofrecen el transfer learning, los modelos ligeros y la integración multimodal para mejorar la precisión y escalabilidad de estas soluciones?

No obstante, aún persisten limitaciones que restringen la aplicación práctica de estos sistemas en el cacao. Los conjuntos de datos disponibles suelen ser reducidos o recolectados bajo condiciones específicas, dificultando la generalización de los modelos a escenarios reales caracterizados por variaciones de iluminación, oclusiones y fondos heterogéneos (Murat & Kiran, 2025; Zahra et al., 2024). Los detectores también pueden presentar sensibilidad frente a la calidad de las imágenes y a la identificación de síntomas incipientes (Paul et al., 2025). De igual manera, las propuestas basadas en Transformers y arquitecturas híbridas muestran un elevado potencial en diferentes cultivos, pero su aplicación al diagnóstico de enfermedades del cacao continúa siendo un campo con oportunidades de desarrollo (Bhowmik et al., 2024; Vallabhajosyula et al., 2024).

Finalmente, tecnologías como los vehículos aéreos no tripulados (UAV), los sensores y los sistemas de análisis espectral están ampliando las posibilidades de la agricultura de precisión, aunque su integración en plantaciones de cacao todavía requiere mayor investigación (Al-Obeidat & Li, 2025; Hafizal et al., 2022). En consecuencia, aunque las evidencias obtenidas en diferentes cultivos resultan prometedoras, persiste la necesidad de organizar y evaluar críticamente las arquitecturas utilizadas específicamente en cacao, así como sus condiciones de aplicación en escenarios reales de producción.

Este estudio tiene como finalidad sintetizar y analizar la literatura científica sobre visión por computadora aplicada a la detección de enfermedades fúngicas en cacao, con énfasis en

arquitecturas de Deep Learning. Con ello, se pretende organizar el conocimiento existente, contrastar los principales enfoques e identificar oportunidades de investigación que contribuyan al desarrollo de sistemas más robustos, eficientes y sostenibles para la agricultura tropical.

2. Metodología

2.1. Diseño de estudio

Esta investigación se planteó como una Revisión Sistemática de Literatura (RSL) bajo las directrices de PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), lo que permitió organizar el proceso de búsqueda y selección de manera transparente y reproducible. El diseño se estructuró siguiendo las fases de identificación, cribado, elegibilidad e inclusión, documentadas en el diagrama de flujo PRISMA, con el fin de reducir sesgos y garantizar la exhaustividad en la selección de los estudios.

Para delimitar con claridad el alcance de la revisión se utilizó además el marco PICOC (Population, Intervention, Comparison, Outcome, Context). En este caso, la población correspondió a cultivos de cacao afectados por enfermedades fúngicas, la intervención al uso de arquitecturas de deep learning aplicadas a visión por computadora, la comparación a los métodos tradicionales de diagnóstico y estudios en cultivos agrícolas afines, los resultados a métricas de desempeño y capacidad de generalización, y el contexto a sistemas agrícolas tropicales en escenarios de campo real y aplicaciones tecnológicas como dispositivos móviles y UAV. La integración de PRISMA y PICOC otorga solidez al diseño metodológico y asegura que el análisis realizado en esta revisión sintetice de manera rigurosa y crítica el estado del arte en técnicas de visión por computadora aplicadas a la detección de enfermedades fúngicas en cacao.

2.2. Criterios de elegibilidad

Los criterios de elegibilidad se definieron con el fin de asegurar que los estudios seleccionados fueran relevantes para el objetivo de analizar el uso de técnicas de deep learning en la detección de enfermedades fúngicas en cultivos de cacao. A continuación, se presentan los criterios de inclusión y exclusión empleados.

Tabla 1. Criterios de inclusión y exclusión

Inclusión	Exclusión
• Uso de Deep Learning: Estudios que empleen CNN, YOLO, U-Net o Visión Transformers en tareas de clasificación, detección o segmentación de enfermedades fúngicas en cacao. • Enfoque en Cacao: Solo estudios relacionados con el cultivo de cacao. • Métricas Cuantitativas: Artículos que reporten precisión, recall, F-score, o mAP. • Publicaciones en inglés o español. • Metodología Reproducible: Descripción detallada de los datasets y métodos de procesamiento.	• Revisiones Teóricas: Se excluyen estudios sin validación experimental directa (metaanálisis, revisiones). • Machine Learning Clásico: Investigación basada únicamente en métodos como SVM o Random Forest. • Enfermedades en Otros Cultivos: Estudio de plagas o enfermedades en cultivos no relacionados con cacao. • Falta de Métricas Cuantitativas: Artículos sin datos de evaluación del rendimiento de los modelos. • Artículos Duplicados o Preliminares: Artículos duplicados o en fase preliminar. • Fuentes No Revisión por Pares: Se descartan tesis, libros, patentes y fuentes no revisadas científicamente.

Nota: Esta tabla define los criterios utilizados para filtrar estudios relevantes en la detección de enfermedades fúngicas del cacao, priorizando aquellos con validación experimental, métricas claras y enfoque aplicado.

2.3. Estrategias de búsqueda

La estrategia de búsqueda se diseñó de manera sistemática con el propósito de identificar estudios relevantes sobre la aplicación de técnicas de visión por computadora y deep learning para la detección de enfermedades fúngicas en el cultivo de cacao. La búsqueda bibliográfica se realizó en tres bases de datos científicas: Web of Science (WoS), ScienceDirect y Scopus, seleccionadas por su cobertura de publicaciones en las áreas de ingeniería, ciencias de la computación y ciencias agrícolas.

La cadena de búsqueda se estructuró mediante términos agrupados en cuatro bloques temáticos: (i) cultivo de interés, (ii) problema fitosanitario, (iii) tecnología empleada y (iv) tarea computacional. Para establecer las relaciones entre los términos se utilizaron los operadores booleanos AND y OR, así como comillas para delimitar frases exactas. La estrategia general de búsqueda fue la siguiente:

TITLE-ABS-KEY ("cacao" OR "theobroma cacao" OR "cocoa")

AND

TITLE-ABS-KEY ("fungal disease" OR "plant disease" OR "black pod"

OR "moniliasis" OR "witches' broom" OR "moniliophthora"

OR "phytophthora")

AND

TITLE-ABS-KEY ("computer vision" OR "deep learning"

OR "machine learning" OR "artificial intelligence")

AND

TITLE-ABS-KEY ("convolutional neural network" OR "cnn"

OR "u-net" OR "yolo" OR "vision transformer" OR "vit")

AND

TITLE-ABS-KEY ("image classification" OR "object detection"

OR "image segmentation" OR "detection" OR "identification"

OR "classification")

Debido a las particularidades de sintaxis y campos de búsqueda de cada base de datos, la ecuación se adaptó según los requerimientos de Web of Science, ScienceDirect y Scopus, manteniendo los mismos bloques conceptuales y términos principales.

La ejecución de la estrategia permitió identificar inicialmente 305 registros, distribuidos en 22 documentos en Web of Science, 145 en ScienceDirect y 138 en Scopus. Los registros recuperados fueron exportados a Mendeley para la gestión bibliográfica y a Parsifal para apoyar el proceso de revisión sistemática, incluyendo la organización de los estudios y la posterior identificación y eliminación de registros duplicados.

2.4. Procesos de selección

Se identificaron 305 artículos en las bases de datos consultadas. Tras eliminar 67 duplicados, quedaron 238 registros únicos para el cribado. En esta fase, se excluyeron 114 artículos por no cumplir los criterios de elegibilidad. Los 124 artículos restantes se sometieron a una evaluación de texto completo. De estos, 71 fueron excluidos por las siguientes razones: 45 por no utilizar arquitecturas de Deep Learning modernas, 23 por carecer de métricas de rendimiento cuantitativas, y 13 por ser estudios de revisión teórica sin experimentación propia. Finalmente,

43 artículos cumplieron todos los criterios y fueron incluidos en la revisión sistemática. Este procedimiento se resume en el diagrama PRISMA (Figura 1).

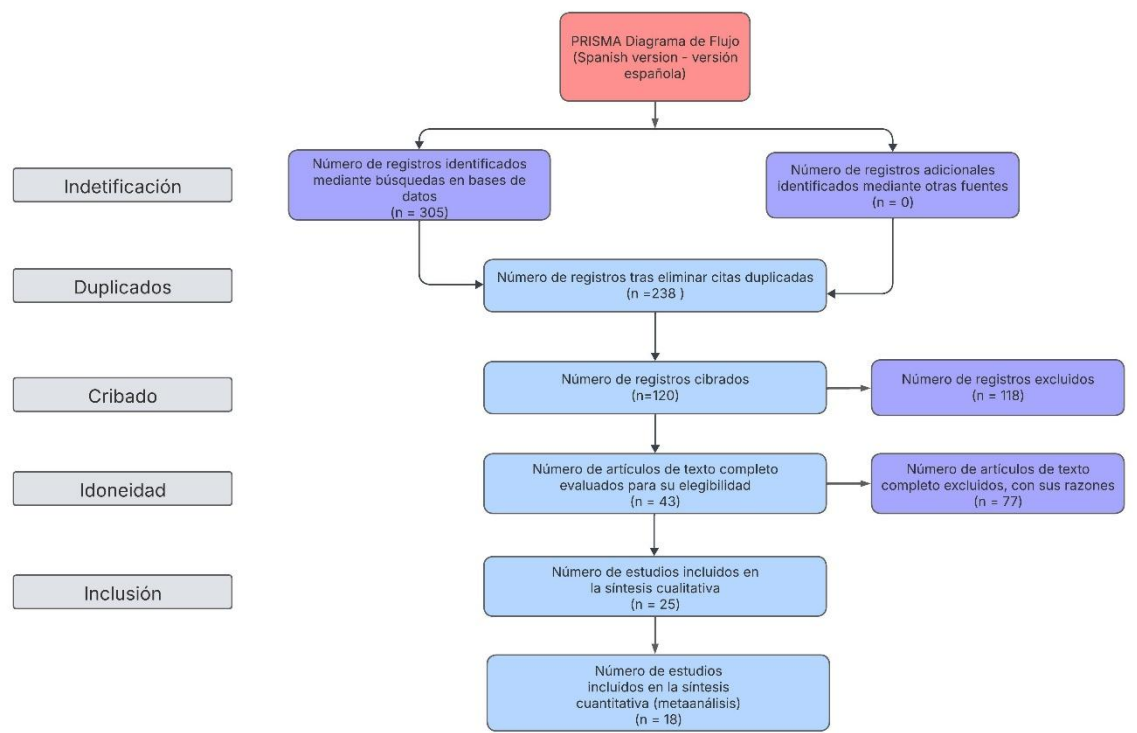


Figura 1. Diagrama de Flujo PRISMA (Spanish version - versión española)

Nota: Proceso de selección de artículos, aplicando PRISMA.

2.5 Evaluación de la calidad

Se aplicó una rúbrica de cinco criterios ponderados: descripción del dataset (CM), reporte de métricas y comparativas (RE), enfoque en enfermedades reales (RA), publicación peer-reviewed (CP) e innovación (CC). Cada estudio se calificó con 1 = Sí, 0.5 = Parcial o 0 = No. Con los pesos de la columna “Puntuación” (CM=1, RE=0.9, RA=0.8, CP=0.6, CC=0.4) se calculó un índice de calidad ponderado y se fijó un umbral de inclusión ≥ 0.70 , como se resume en la tabla N°02.

Tabla 2. Criterios de calidad

Criterio	Abreviación	Puntuación
Descripción del dataset de cultivos con enfermedades fúngicas.	CM	1
Uso de métricas estándar y comparación con otros métodos.	RE	1
Enfoque en enfermedades fúngicas de cultivos reales.	RA	1
Revista/conferencia indexada y peer-reviewed.	CP	0.5
Propuesta innovadora frente a estudios previos.	CC	0.5

Nota: La evaluación de la calidad se realizó mediante una rúbrica de cinco criterios ponderados, asignando puntajes (1 = Sí, 0.5 = Parcial, 0 = No).

3. Desarrollo

Los estudios revisados abarcan una amplia variedad de enfoques aplicados a la detección de enfermedades en plantas. Se identificaron aportes específicos en cacao, desarrollos en otros cultivos con potencial transferible y propuestas transversales relacionadas con eficiencia, explicabilidad y multimodalidad.

3.1. Resultados del cacao

Los estudios enfocados específicamente en cacao evidenciaron avances en la detección de enfermedades, la clasificación de estados fenológicos y el desarrollo de herramientas móviles de apoyo. Nyarko et al. (2024) propusieron un modelo YOLOv5m optimizado mediante Swin Transformer, K-means++ y la función de pérdida EIou para la detección de enfermedades del cacao, alcanzando 96 % de precisión, 94 % de recall y 92 % de mAP, lo que evidencia el potencial de las arquitecturas YOLO optimizadas para tareas de detección automatizada.

En el ámbito de las aplicaciones móviles, Kumi et al. (2022) desarrollaron Cocoa Companion, una aplicación para teléfonos inteligentes basada en deep learning orientada a la detección de enfermedades del cacao. Este enfoque evidencia la posibilidad de trasladar los modelos de visión por computadora hacia herramientas accesibles para su utilización fuera de entornos computacionales convencionales.

Respecto a la clasificación de estados fenológicos, Borromeo et al. (2025) emplearon redes neuronales convolucionales para la clasificación del grado de madurez de las mazorcas de cacao, mostrando la aplicabilidad de estas arquitecturas en tareas de reconocimiento visual asociadas al cultivo. Por otra parte, Hafizal et al. (2022) implementaron un sistema experto basado en forward chaining para identificar deficiencia de potasio en plantas de cacao, ampliando el espectro de herramientas computacionales utilizadas para apoyar el diagnóstico del estado del cultivo.

Desde una perspectiva más amplia, Sykes et al. (2024) analizaron las aplicaciones de visión por computadora en patología vegetal utilizando ejemplos procedentes de la agricultura del cacao. Su trabajo constituye un referente para comprender tanto las posibilidades de estas tecnologías como los desafíos asociados con su implementación en escenarios agrícolas.

En conjunto, la evidencia específica disponible para cacao muestra una evolución desde sistemas basados en reglas y arquitecturas CNN hacia modelos de detección más avanzados, particularmente variantes optimizadas de YOLO. No obstante, el número limitado de investigaciones directamente centradas en enfermedades del cacao evidencia una brecha en la disponibilidad de estudios experimentales y conjuntos de datos específicos del cultivo, aspecto que justifica complementar el análisis con investigaciones desarrolladas en otros cultivos potencialmente transferibles al contexto cacaotero.

3.2. Resultados en otros cultivos (transferibles a cacao)

3.2.1. Modelos CNN y YOLO

En tomate, Ruhad et al. (2025) evaluaron modelos de detección de objetos para la identificación de enfermedades foliares utilizando un conjunto de datos obtenido en condiciones reales. Los resultados reportados en el manuscrito muestran que YOLOv9 con CSPDarknet-53 alcanzó un F1 de 97.86 %, evidenciando un desempeño favorable frente a otras arquitecturas de detección consideradas.

También en tomate, Yulita et al. (2023) desarrollaron una aplicación móvil para la detección de enfermedades foliares mediante una arquitectura DenseNet, demostrando la posibilidad de implementar redes convolucionales en aplicaciones orientadas al diagnóstico automatizado. Complementariamente, Mashamba et al. (2024) utilizaron una CNN para la detección de la mancha bacteriana en hojas de tomate, reforzando la aplicabilidad de estas arquitecturas en tareas de reconocimiento de enfermedades foliares.

En sandía, Madadum et al. (2025) propusieron un enfoque basado en YOLO con aumento de datos mediante SAM para optimizar la detección de enfermedades foliares. En el manuscrito se reportan valores de mAP@0.5 de 0.92 y mAP@0.95 de 0.71, lo que evidencia el potencial de este enfoque para la identificación automatizada de síntomas en escenarios agrícolas.

En maíz, Nkuna et al. (2025) compararon modelos CNN y ResNet50 para la identificación de enfermedades foliares a partir de imágenes RGB. De acuerdo con los resultados consignados en el manuscrito, ResNet50 alcanzó una exactitud de 78.76 %, superior al 71.01 % obtenido mediante una CNN entrenada desde cero, lo que respalda el potencial de arquitecturas residuales para este tipo de tareas.

En papaya, Bacus y Linsangan (2022) desarrollaron una solución para la detección e identificación de hojas de Carica papaya mediante Android, constituyendo evidencia adicional sobre la posibilidad de trasladar modelos de reconocimiento visual hacia dispositivos móviles.

En conjunto, los estudios analizados muestran que las arquitecturas CNN y los modelos de la familia YOLO constituyen alternativas relevantes para la identificación y detección automatizada de enfermedades en diferentes cultivos. Los resultados obtenidos en tomate, sandía, maíz y papaya aportan evidencia potencialmente transferible al cacao; sin embargo, dicha transferencia debe interpretarse con cautela debido a las diferencias entre cultivos, síntomas, condiciones de adquisición de imágenes y conjuntos de datos empleados.

3.2.2. Modelos basados en Transformers

Los Vision Transformers (ViT) han mostrado un desempeño elevado en tareas de clasificación de enfermedades vegetales. Bhowmik et al. (2024) desarrollaron un ViT personalizado para la detección y clasificación de enfermedades foliares en Java plum, evidenciando la capacidad de estas arquitecturas para abordar tareas de reconocimiento visual en contextos agrícolas.

En manzana, Huang et al. (2025) propusieron EConv-ViT, una arquitectura que integra ConvNeXt y Transformer con el propósito de mejorar la generalización en la clasificación de enfermedades foliares. Este enfoque híbrido evidencia el potencial de combinar características convolucionales con mecanismos de atención para fortalecer la discriminación de patrones asociados con enfermedades vegetales.

Por su parte, Vallabhajosyula et al. (2024) desarrollaron un marco jerárquico basado en Residual Vision Transformer para la detección de enfermedades en hojas de plantas, mostrando la aplicabilidad de los Transformers como alternativa a las arquitecturas convolucionales convencionales.

En escenarios de detección de enfermedades en cultivos, Tong et al. (2025) propusieron ToT-Net, un modelo orientado a la detección generalizada y en tiempo real que integra aprendizaje a nivel de tarea y mecanismos Transformer multiescala. En el manuscrito se reporta un desempeño de 94.6 % de mAP, F1 = 0.905 y 102.3 FPS, resultados que respaldan su potencial para aplicaciones que requieren procesamiento en tiempo real.

En conjunto, estos estudios muestran que las arquitecturas basadas en Transformers y los modelos híbridos CNN-Transformer representan una línea prometedora para el diagnóstico automatizado de enfermedades vegetales, debido a su capacidad para modelar relaciones globales en las imágenes y favorecer la generalización. Sin embargo, su transferencia al diagnóstico de enfermedades fúngicas del cacao todavía requiere mayor validación experimental, especialmente bajo condiciones reales de campo.

3.2.3. Enfoques multimodales y UAVs

En el ámbito de la agricultura de precisión, Al-Obeidat y Li (2025) desarrollaron un pipeline multimodal basado en vehículos aéreos no tripulados (UAV), que integró YOLO para la detección aérea y DeiT para la clasificación de enfermedades. En el manuscrito se reporta que este enfoque

alcanzó 95.4 % de mAP y 99.45 % de exactitud, evidenciando el potencial de combinar plataformas aéreas y arquitecturas de deep learning para el monitoreo automatizado de cultivos.

Por otro lado, Zahra et al. (2024) analizaron los avances y aplicaciones de la espectroscopía de imágenes, tecnología que amplía las posibilidades de caracterización mediante información espectral. No obstante, la referencia incluida en el manuscrito es de carácter general y no permite atribuir directamente a este estudio una mejora significativa en la identificación de patógenos ni afirmar específicamente que el costo de los sensores constituye una barrera, por lo que conviene evitar esas afirmaciones tal como estaban redactadas.

3.3. Resultados transversales

3.3.1. Explicabilidad (XAI)

Las técnicas de inteligencia artificial explicable (XAI) han adquirido relevancia en el diagnóstico automatizado de enfermedades vegetales al permitir interpretar los criterios utilizados por los modelos durante la clasificación. En este contexto, herramientas como Grad-CAM, LIME y SHAP permiten identificar las características o regiones de las imágenes que ejercen mayor influencia sobre las predicciones, contribuyendo a mejorar la interpretabilidad de los modelos de deep learning. En particular, Porfírio et al. (2025) analizaron la aplicación de enfoques XAI en el reconocimiento de enfermedades vegetales mediante deep learning, destacando su utilidad para comprender las decisiones generadas por estos sistemas.

Asimismo, la incorporación de mecanismos de explicabilidad adquiere especial importancia en herramientas destinadas a usuarios finales, debido a que no solo interesa obtener una clasificación precisa, sino también proporcionar elementos que permitan comprender el fundamento de la predicción. En este sentido, los sistemas de diagnóstico basados en deep learning pueden beneficiarse de estrategias que incrementen la transparencia de sus resultados y faciliten su interpretación en contextos agrícolas.

3.3.2. Optimización y eficiencia

El desarrollo de modelos computacionalmente eficientes constituye una línea relevante para facilitar la implementación de sistemas de diagnóstico automatizado en dispositivos con recursos limitados. Yao et al. (2025) propusieron WTAD-YOLO, un modelo ligero para la detección de enfermedades foliares que, con 2.32 millones de parámetros, alcanzó un $mAP@0.5 = 0.917$, mostrando un equilibrio favorable entre complejidad arquitectónica y capacidad de detección.

En tareas de segmentación, Ibtehz y Rahman (2020) desarrollaron MultiResUNet, una modificación de U-Net que incorpora bloques multirresolución para mejorar la representación de características a diferentes escalas. Aunque esta arquitectura fue evaluada originalmente en segmentación de imágenes biomédicas, sus principios arquitectónicos constituyen un antecedente relevante para el diseño de modelos de segmentación capaces de representar estructuras visuales de distinta escala.

Asimismo, diferentes modificaciones de U-Net muestran cómo la optimización arquitectónica puede orientarse a tareas específicas. Li et al. (2026) propusieron RCPU-Net, una arquitectura de segmentación multiescala basada en U-Net, mientras que Kazi et al. (2025) desarrollaron VCNet, un marco de deep learning para la clasificación multiclase de enfermedades del arroz mediante extracción profunda de características y optimización con algoritmo genético. Estos trabajos evidencian diferentes estrategias de optimización de arquitecturas profundas, aunque sus resultados no deben interpretarse directamente como evidencia de portabilidad en dispositivos agrícolas de campo.

3.3. Transfer learning y pocos datos

El transfer learning constituye una estrategia relevante para abordar escenarios agrícolas caracterizados por la disponibilidad limitada de datos. En cacao, Borromeo et al. (2025) emplearon transferencia de aprendizaje con VGG16 para la clasificación de la madurez de

mazorcas, evidenciando la utilidad de modelos preentrenados en tareas de reconocimiento visual.

En maíz, Nkuna et al. (2025) evaluaron ResNet50 mediante transferencia de aprendizaje, alcanzando una exactitud de 78.76 %, superior al 71.01 % obtenido mediante una CNN entrenada desde cero. Estos resultados respaldan el potencial del aprovechamiento de representaciones previamente aprendidas cuando se dispone de conjuntos de datos agrícolas relativamente limitados.

Tabla 3. Comparaciones de modelos

Tarea	Modelos	Dataset	Metricas	Precisión
enfermedades foliares	yolov5m + swint + elou	1000 imágenes	precisión / recall / map	96 % / 94% / 92%
diagnostico móvil	sdd mobilenet v2	app android	confidence	> 80%
diagnostico móvil	sdd-resnet50 v1-fpn	app android	recall	52.9% / 59.5 %
madurez fenologica	vgg16(tl)	dataset reducido	exactitud	91.2%
deficiencia nutricional	sistema experto	reglas (forward chaining)	diagnostico	88%

Nota. Se presentan los principales modelos aplicados al diagnóstico en cacao, detallando su dataset, métricas empleadas y nivel de desempeño alcanzado.

Tabla 4. Modelos en otros cultivos

Cultivo	Modelo	Dataset/Escenario	Metricas	Desempeño
Tomate	Yolov9 (Cspdarknet53)	Benchmark	F1	97.86%
Tomate	Densenet Movil	Campo Real	Precisión	> 90%
Sandía	Yolov9 + Sam	Campo Real	Map50/Map95	0.92 / 0.71
Maíz	Resnet(Tl)	1500 Imagenes	Excactitud	78.76%
Arroz	CNN Clásicas	Estrés Follar	Precisión	= 90%

Nota. Se muestran los resultados de diferentes modelos aplicados en cultivos diversos, comparando arquitecturas, métricas y rendimientos obtenidos

4. Discusión

4.1. Limitaciones del Estudio

La escasez de estudios aplicados específicamente al cacao hizo necesario incorporar investigaciones desarrolladas en otros cultivos con problemas fitosanitarios comparables. Esta decisión permitió ampliar el análisis de arquitecturas y estrategias computacionales, aunque introduce una limitación de transferibilidad, debido a que las características visuales de los síntomas, las condiciones ambientales y los protocolos de adquisición de imágenes varían entre especies. En este sentido, la literatura sobre visión por computadora aplicada a la patología vegetal evidencia avances importantes, pero también señala desafíos para trasladar el desempeño obtenido experimentalmente hacia condiciones agrícolas reales (Sykes et al., 2024; Sapkota & Karkee, 2026; Leite et al., 2024).

En relación con las arquitecturas evaluadas, los resultados muestran un equilibrio particularmente favorable en los modelos de la familia YOLO y en las CNN optimizadas. Yao et al. (2025) desarrollaron WTAD-YOLO, una arquitectura ligera con 2.32 millones de parámetros y un mAP@0.5 de 0.917, evidenciando que la reducción de la complejidad arquitectónica puede mantenerse compatible con un elevado desempeño de detección. En contraste, los modelos basados en Transformers e híbridos CNN-Transformer ofrecen mecanismos más sofisticados

para capturar información global. Huang et al. (2025) propusieron EConv-ViT mediante la integración de ConvNeXt y Transformer para la clasificación de enfermedades foliares de manzana. Sin embargo, las diferencias entre ambos trabajos deben interpretarse con cautela, dado que fueron evaluados en cultivos, conjuntos de datos y tareas diferentes, por lo que sus métricas no constituyen una comparación directa entre arquitecturas.

Esta heterogeneidad también se observa entre estudios basados en YOLO. En tomate, Ruhad et al. (2025) evaluaron modelos de detección sobre un conjunto de datos obtenido en condiciones reales, mientras que Madadum et al. (2025) emplearon un enfoque basado en YOLO y aumento de datos mediante SAM para enfermedades foliares de sandía. Las diferencias de desempeño observadas entre estudios indican que la eficacia de una arquitectura no depende exclusivamente del modelo, sino también del cultivo, la naturaleza y severidad de los síntomas, la composición del conjunto de datos y las condiciones de adquisición de las imágenes. En consecuencia, la evidencia revisada no permite establecer una arquitectura universalmente superior para todos los escenarios agrícolas.

Una de las principales brechas identificadas corresponde a la disponibilidad y representatividad de los conjuntos de datos. La utilización de imágenes obtenidas bajo condiciones controladas puede favorecer el desempeño de los modelos, pero limita la posibilidad de generalizar sus resultados a escenarios caracterizados por variaciones de iluminación, fondos complejos, oclusiones y diferentes etapas de desarrollo de la enfermedad. Por ello, la validación en condiciones reales de campo constituye un aspecto fundamental para determinar la utilidad práctica de los sistemas de visión por computadora. Las revisiones de Jackulin y Murugavalli (2022) y Sykes et al. (2024) permiten situar esta problemática dentro de los desafíos generales de la detección automatizada de enfermedades vegetales.

Frente a la disponibilidad limitada de imágenes, el transfer learning constituye una de las estrategias con evidencia más consistente dentro de los estudios analizados. En cacao, Borromeo et al. (2025) utilizaron VGG16 con transferencia de aprendizaje para la clasificación de la madurez de mazorcas, mientras que Nkuna et al. (2025) evaluaron ResNet50 en enfermedades foliares de maíz, alcanzando una exactitud de 78.76 % frente al 71.01 % de una CNN entrenada desde cero. Estos hallazgos respaldan el aprovechamiento de representaciones previamente aprendidas cuando no se dispone de conjuntos de datos suficientemente amplios para entrenar modelos profundos desde cero.

Otro componente relevante es la inteligencia artificial explicable (XAI). Un modelo destinado al diagnóstico agrícola no solo debe alcanzar un desempeño adecuado, sino proporcionar elementos que permitan interpretar las características visuales asociadas con sus predicciones. Porfirio et al. (2025) exploraron técnicas de XAI aplicadas a la clasificación de enfermedades vegetales, evidenciando la importancia creciente de la interpretabilidad dentro de la agricultura digital. Esta dimensión adquiere particular relevancia en herramientas orientadas a técnicos y productores, donde la comprensión de la predicción puede complementar las métricas convencionales de exactitud, sensibilidad o precisión.

Asimismo, la literatura muestra una tendencia hacia la integración de diferentes fuentes de información y plataformas de adquisición. Al-Obeidat y Li (2025) propusieron un pipeline multimodal basado en UAV que combina YOLO para detección aérea y DeiT para clasificación, mostrando el potencial de integrar adquisición remota y aprendizaje profundo en agricultura de precisión. Paralelamente, los avances en espectroscopía de imágenes analizados por Zahra et al. (2024) amplían las posibilidades de incorporar información espectral al análisis automatizado. Sin embargo, la evidencia incluida en esta revisión todavía resulta insuficiente para establecer la viabilidad económica y operativa de estas tecnologías específicamente en pequeños productores de cacao.

En conjunto, los hallazgos permiten responder las preguntas de investigación planteadas. En cuanto a las arquitecturas más efectivas, los modelos de la familia YOLO y las CNN con transferencia de aprendizaje presentan un balance favorable entre desempeño y complejidad,

mientras que los Transformers y las arquitecturas híbridas constituyen alternativas prometedoras para mejorar la representación de características complejas. Respecto a las limitaciones actuales, destacan la escasez de conjuntos de datos específicos de cacao, la heterogeneidad metodológica entre estudios, la limitada validación bajo condiciones reales y la todavía reducida incorporación de mecanismos de explicabilidad. Finalmente, como tendencias emergentes, se identifican el desarrollo de arquitecturas ligeras, la transferencia de aprendizaje, los modelos híbridos CNN-Transformer, la XAI y la integración de UAV y datos multimodales.

A diferencia de revisiones generales sobre aprendizaje profundo y visión por computadora aplicada a enfermedades vegetales, la presente revisión delimita su análisis hacia las enfermedades fúngicas del cacao y utiliza el marco PRISMA-PICOC para estructurar el proceso de identificación, selección y análisis de la evidencia. Este enfoque permite no solo examinar el desempeño reportado por diferentes arquitecturas, sino también contrastar aspectos transversales relacionados con eficiencia computacional, explicabilidad, multimodalidad y posibilidades de implementación. No obstante, debido a la limitada cantidad de estudios directamente centrados en cacao, parte de la evidencia procede de otros cultivos y debe considerarse como evidencia potencialmente transferible, no como demostración directa de desempeño en cacao.

Finalmente, la síntesis realizada evidencia una brecha entre el desempeño experimental y las necesidades de implementación en escenarios agrícolas reales. Aunque las CNN continúan ofreciendo ventajas asociadas con su madurez tecnológica y disponibilidad de arquitecturas preentrenadas, los Transformers y modelos híbridos amplían las posibilidades de representación y generalización. Sin embargo, para el contexto del cacao, la prioridad no debería limitarse a incrementar las métricas de desempeño, sino avanzar hacia conjuntos de datos representativos de diferentes condiciones agroclimáticas, etapas de la enfermedad y escenarios regionales. Esta necesidad resulta especialmente relevante para orientar futuros desarrollos hacia modelos que puedan ser evaluados y posteriormente desplegados en entornos tropicales de producción.

5. Conclusiones

La revisión sistemática permitió identificar avances sólidos en la detección de enfermedades del cacao mediante aprendizaje profundo. Se comprobó que YOLO optimizado y CNN con transfer learning constituyen las arquitecturas más efectivas en condiciones de campo, mientras que los Transformers ofrecen un mayor potencial de generalización, aunque limitados por su complejidad.

La principal barrera detectada es la escasez de datasets representativos, acompañada de la falta de explicabilidad y la baja validación en escenarios reales. Estos factores explican la disparidad en métricas entre estudios controlados y pruebas en campo abierto.

Para el periodo 2025–2030, la agenda de investigación debe priorizar: (i) la creación de un benchmark público de cacao con al menos 10,000 imágenes en campo real, (ii) el desarrollo de modelos ligeros optimizados para edge computing, y (iii) la incorporación sistemática de explicabilidad y multimodalidad en las aplicaciones móviles. Con estos avances, los logros académicos podrán transformarse en herramientas prácticas de impacto directo en la productividad y sostenibilidad de la cadena cacaotera.

Financiamiento

Ninguno.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Contribución de autores

Yomer Ysidro Santos: Conceptualización, redacción del borrador original, introducción, resultados, discusión y conclusiones.

Andy Roel Vasquez Bueno: Diseño del estudio, criterios de elegibilidad, evaluación de calidad y organización de referencias bibliográficas.

Isait Camizán Ojeda: Estrategias de búsqueda, procesos de selección y elaboración de gráficos y figuras.

Los tres autores colaboraron conjuntamente en la revisión de artículos mediante las herramientas Mendeley y Parsifal, contribuyendo a la sistematización de datos y validación del análisis final.

Referencias bibliográficas

Al-Obeidat, N., & Li, Z. (2025). A multimodal UAV-based pipeline for precision agriculture: Aerial stress detection with YOLO and high-fidelity disease classification using DeiT. *Procedia Computer Science*, 921–926. <https://doi.org/10.1016/j.procs.2025.03.118>

Bacus, J. A., & Linsangan, N. B. (2022). Detection and identification with analysis of *Carica papaya* leaf using Android. *Journal of Advances in Information Technology*, 13(2), 162–166. <https://doi.org/10.12720/jait.13.2.162-166>

Bhowmik, A. C., Ahad, M. T., Emon, Y. R., Ahmed, F., Song, B., & Li, Y. (2024). A customised vision transformer for accurate detection and classification of Java plum leaf disease. *Smart Agricultural Technology*, 8, 100500. <https://doi.org/10.1016/j.atech.2024.100500>

Borromeo, K. O., Bonotan, M. R., Llagas, A. P., Juan, J. M. O., Sanchez, P. D. C., & Berro, N. K. A. (2025). Classification of cacao pods ripeness using convolutional neural networks. *Procedia Computer Science*, 257, 1197–1204. <https://doi.org/10.1016/j.procs.2025.03.160>

Hafizal, M. T., et al. (2022). Implementation of expert systems in potassium deficiency in cocoa plants using forward chaining method. *Procedia Computer Science*, 136–143. <https://doi.org/10.1016/j.procs.2022.12.120>

Huang, X., et al. (2025). EConv-ViT: A strongly generalized apple leaf disease classification model based on the fusion of ConvNeXt and Transformer. *Information Processing in Agriculture*. <https://doi.org/10.1016/j.inpa.2025.03.001>

Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121, 74–87. <https://doi.org/10.1016/j.neunet.2019.08.025>

Jackulin, C., & Murugavalli, S. (2022). A comprehensive review on detection of plant disease using machine learning and deep learning approaches. *Measurement: Sensors*, 24, 100441. <https://doi.org/10.1016/j.measen.2022.100441>

Kazi, S. S., Palkar, B., & Mishra, D. (2025). VCNet: Optimized deep learning framework with deep feature extraction and genetic algorithm for multiclass rice crop disease detection. *MethodsX*, 15, 103551. <https://doi.org/10.1016/j.mex.2025.103551>

Kumi, S., Kelly, D., Woodstuff, J., Lomotey, R. K., Orji, R., & Deters, R. (2022). Cocoa Companion: Deep learning-based smartphone application for cocoa disease detection. *Procedia Computer Science*, 203, 87–94. <https://doi.org/10.1016/j.procs.2022.07.013>

- Leite, D., Brito, A., & Faccioli, G. (2024). Advancements and outlooks in utilizing convolutional neural networks for plant disease severity assessment: A comprehensive review. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2024.100573>
- Li, N., Zhang, A., Han, H., & Duan, Y. (2026). RCPU-Net: A multi-scale multi-object segmentation model for coal and gangue under uneven lighting based on improved U-Net. *Digital Signal Processing*, 168, 105484. <https://doi.org/10.1016/j.dsp.2025.105484>
- Madadum, H., Nasir, F. E., & Haruehansapong, K. (2025). Optimizing watermelon leaf disease detection using SAM-based augmentation with YOLO for practical agricultural solutions. *Smart Agricultural Technology*, 12, 101326. <https://doi.org/10.1016/j.atech.2025.101326>
- Mashamba, M. M., Telukdarie, A., Munien, I., Onkonkwo, U., & Vermeulen, A. (2024). Detection of bacterial spot disease on tomato leaves using a convolutional neural network (CNN). *Procedia Computer Science*, 602–609. <https://doi.org/10.1016/j.procs.2024.05.145>
- Nkuna, B. L., Abutaleb, K., Chirima, J. G., Newete, S. W., van der Walt, A. J., & Nyamugama, A. (2025). Identification of maize leaf diseases using red, green, blue-based images with convolutional neural network (CNN) and residual network (ResNet50) models. *Smart Agricultural Technology*, 12, 101226. <https://doi.org/10.1016/j.atech.2025.101226>
- Nyarko, B. N. E., Wu, W., Zhou, Z. J., & Mohd, M. J. (2024). Improved YOLOv5m model based on Swin Transformer, K-means++, and Efficient Intersection over Union (EIoU) loss function for cocoa tree (*Theobroma cacao*) disease detection. *Journal of Plant Protection Research*, 64(3), 265–274. <https://doi.org/10.24425/jppr.2024.151253>
- Porfírio, R. P., Madeira, R. N., & Santos, P. A. (2025). Exploring explainable AI techniques for plant disease classification in digital agriculture. *Procedia Computer Science*, 175–182. <https://doi.org/10.1016/j.procs.2025.07.170>
- Ruhad, F. M., Fahim, M., Hossain, M. S., Monir, M. F., Islam, A., & Amin, M. A. (2025). Beyond classification: Benchmarking object detection models for efficient tomato leaf disease identification on a real-world dataset. *Smart Agricultural Technology*, 12, 101336. <https://doi.org/10.1016/j.atech.2025.101336>
- Sapkota, R., & Karkee, M. (2026). Object detection with multimodal large vision-language models: An in-depth review. *Information Fusion*, 126, 103575. <https://doi.org/10.1016/j.inffus.2025.103575>
- Sykes, J. R., Denby, K. J., & Franks, D. W. (2024). Computer vision for plant pathology: A review with examples from cocoa agriculture. *Applications in Plant Sciences*, 12(2), e11559. <https://doi.org/10.1002/aps3.11559>
- Tong, J., Zhang, L., Tian, J., Yu, Q., & Lang, C. (2025). ToT-Net: A generalized and real-time crop disease detection framework via task-level meta-learning and lightweight multi-scale transformer. *Smart Agricultural Technology*, 12, 101249. <https://doi.org/10.1016/j.atech.2025.101249>
- Vallabhajosyula, S., Sistla, V., & Kolli, V. K. K. (2024). A novel hierarchical framework for plant leaf disease detection using residual vision transformer. *Heliyon*, 10(9), e29912. <https://doi.org/10.1016/j.heliyon.2024.e29912>
- Yao, J., Li, Y., Xia, Z., Nie, P., Li, X., & Li, Z. (2025). WTAD-YOLO: A lightweight tomato leaf disease detection model based on YOLO11. *Smart Agricultural Technology*, 12, 101349. <https://doi.org/10.1016/j.atech.2025.101349>

Yulita, I. N., Amri, N. A., & Hidayat, A. (2023). Mobile application for tomato plant leaf disease detection using a dense convolutional network architecture. *Computation*, 11(2), 20.

<https://doi.org/10.3390/computation11020020>

Zahra, A., et al. (2024). Current advances in imaging spectroscopy and its state-of-the-art applications. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2023.122172>