

ORIGINAL

## Deep Learning-Driven Computer Vision for Early and Automatic Detection of Cacao Pests and Diseases

### Visión por computadora impulsada por aprendizaje profundo para la detección temprana y automática de plagas y enfermedades del cacao

Jorge Raúl Navarro-Cabrera<sup>1</sup>  , José Guillermo Beraún-Barrantes<sup>1</sup>  , Ángel Cárdenas-García<sup>2</sup>  , Carlos Mauricio Lozano-Carranza<sup>3</sup>  

<sup>1</sup>Universidad de Huánuco. Huánuco. Perú.

<sup>2</sup>Universidad Nacional de San Martín, Tarapoto. Perú.

<sup>3</sup>TUSAN Ingenieros Consultores, Tarapoto. Perú.

**Citar como:** Navarro-Cabrera JR, Beraún-Barrantes JG, Cárdenas-García Ángel, Lozano-Carranza CM. Deep Learning-Driven Computer Vision for Early and Automatic Detection of Cacao Pests and Diseases. Salud, Ciencia y Tecnología - Serie de Conferencias. 2025; 4:1762. <https://doi.org/10.56294/sctconf20251762>

Enviado: 12-07-2025

Revisado: 09-09-2025

Aceptado: 24-11-2025

Publicado: 25-11-2025

Editor: Dr. William Castillo-González 

Autor para la correspondencia: Jorge Raúl Navarro-Cabrera 

#### ABSTRACT

**Introduction:** deep learning (DL)-based computer vision has emerged as a promising tool for precision agriculture, particularly for detecting crop diseases and pests automatically. This study evaluated the comparative performance of three state-of-the-art DL architectures for automatic identification of cacao pests and diseases using image analysis.

**Method:** a reproducible pipeline was implemented, encompassing image preprocessing, stratified cross-validation, and inferential statistics through repeated-measures ANOVA. The dataset comprised 4,390 images divided into three highly unbalanced classes: Healthy, Black Pod Rot, and Pod Borer. The architectures—ResNet50, EfficientNet-B0, and ViT-B/16—were fully fine-tuned using the AdamW optimizer, early stopping, and a dynamic learning-rate scheduler.

**Results:** all models achieved mean macro-F1 scores above 0,96, with no statistically significant differences observed among them ( $F = 0,278$ ,  $p = 0,7645$ ). Training curves showed rapid convergence and inter-fold stability, indicating consistent generalization without overfitting.

**Conclusions:** performance outcomes suggest that the effectiveness of the detection system relies more on pipeline design and class-balance management than on the specific DL architecture used. The findings contribute to developing reproducible, efficient intelligent systems for cacao phytosanitary monitoring and support the integration of artificial intelligence into precision agriculture practices.

**Keywords:** Agricultural Informatics; Automated Classification; Image Preprocessing; Model Evaluation; Stratified Validation.

#### RESUMEN

**Introducción:** el aprendizaje profundo (deep learning, DL) aplicado a la visión por computadora se ha consolidado como una herramienta prometedora en la agricultura de precisión, especialmente para la detección automática de plagas y enfermedades en cultivos. El presente estudio evaluó el desempeño comparativo de tres arquitecturas avanzadas de DL para la identificación automática de plagas y enfermedades del cacao mediante análisis de imágenes.

**Método:** se implementó un pipeline reproducible que incluyó preprocesamiento de imágenes, validación cruzada estratificada y análisis estadístico inferencial mediante ANOVA de medidas repetidas. El conjunto

de datos comprendió 4 390 imágenes distribuidas en tres clases altamente desbalanceadas: Healthy, Black Pod Rot y Pod Borer. Las arquitecturas ResNet50, EfficientNet-B0 y ViT-B/16 fueron ajustadas completamente (fine-tuning total) y entrenadas con el optimizador AdamW, parada temprana y programador dinámico de tasa de aprendizaje.

**Resultados:** todos los modelos alcanzaron valores medios de F1 macro superiores a 0,96, sin diferencias estadísticamente significativas entre ellos ( $F = 0,278$ ,  $p = 0,7645$ ). Las curvas de entrenamiento mostraron una convergencia rápida y estabilidad entre pliegues, sin evidencias de sobreajuste.

**Conclusiones:** los resultados indican que la eficacia del sistema depende en mayor medida del diseño del pipeline y del control del desbalance de clases que del tipo de arquitectura empleada. Los hallazgos aportan a la implementación de sistemas inteligentes reproducibles y eficientes para el monitoreo fitosanitario del cacao, promoviendo la adopción de soluciones basadas en inteligencia artificial en la agricultura de precisión.

**Palabras clave:** Agroinformática; Clasificación Automática; Evaluación de Modelos; Preprocesamiento de Imágenes; Validación Estratificada.

## INTRODUCCIÓN

El cacao (*Theobroma cacao* L.) es uno de los cultivos agrícolas más importantes a nivel mundial, con un papel clave en la economía de los países tropicales y en la generación de ingresos para millones de pequeños productores.<sup>(1,2,3,4)</sup> Su grano fermentado y seco constituye la materia prima esencial para la industria del chocolate y diversos productos de confitería.<sup>(5,6)</sup> Sin embargo, la producción de cacao enfrenta serias amenazas derivadas de enfermedades fúngicas y plagas que reducen drásticamente el rendimiento y comprometen la calidad del grano, afectando la competitividad global de la cadena de valor.<sup>(7,8,9)</sup> Entre las patologías más relevantes se encuentran la Black Pod Rot, causada por *Phytophthora* spp., y el Pod Borer, ambas responsables de pérdidas económicas significativas en zonas de producción tropical.<sup>(10,11,12)</sup>

Tradicionalmente, la detección de estas afecciones se ha basado en inspecciones visuales realizadas por agricultores o especialistas, un procedimiento subjetivo que depende de la experiencia individual y de la interpretación morfológica de los síntomas.<sup>(13,14)</sup> Este enfoque, además de ser lento y costoso, presenta una alta variabilidad en la precisión del diagnóstico, lo que retrasa la aplicación de medidas de control efectivas y aumenta la propagación de la enfermedad. En consecuencia, la industria agrícola demanda métodos más confiables, rápidos y reproducibles para la identificación temprana de plagas y enfermedades, que permitan optimizar los procesos de manejo y mejorar la productividad de los cultivos.<sup>(15,16,17)</sup>

En este contexto, los avances en visión por computadora y aprendizaje automático (Machine Learning, ML) han abierto nuevas oportunidades para la automatización de tareas de diagnóstico en la agricultura de precisión.<sup>(18,19)</sup> Estas tecnologías permiten analizar grandes volúmenes de datos visuales mediante algoritmos capaces de reconocer patrones complejos en imágenes de hojas, frutos o tallos.<sup>(20,21)</sup> Los sistemas de visión artificial (SVA), al combinar hardware óptico y software de procesamiento digital de imágenes, se han consolidado como herramientas no destructivas, rápidas y rentables para la caracterización y clasificación de productos agrícolas, logrando niveles de exactitud comparables e incluso superiores a los obtenidos por expertos humanos.<sup>(22,23,24)</sup>

El aprendizaje profundo (Deep Learning, DL), como evolución del ML, ha revolucionado el campo de la visión por computadora al introducir modelos capaces de aprender representaciones jerárquicas directamente a partir de los datos.<sup>(25,26)</sup> Estos enfoques permiten extraer de manera automática características relevantes de las imágenes, eliminando la necesidad de una ingeniería manual de atributos y aumentando la capacidad de generalización en tareas complejas.<sup>(27)</sup> Gracias a su arquitectura multinivel, el DL ha demostrado un desempeño sobresaliente en la detección, segmentación y clasificación de patrones visuales, consolidándose como el paradigma dominante en aplicaciones de diagnóstico agrícola, control de calidad y monitoreo automatizado de cultivos.<sup>(28,29,30)</sup>

No obstante, la literatura actual revela limitaciones importantes. La mayoría de los estudios se centra en una sola arquitectura o evalúa modelos sin aplicar métodos estadísticos que permitan comparar su rendimiento con significancia inferencial. Esta falta de análisis comparativo riguroso impide identificar con claridad qué arquitecturas son más estables, precisas y eficientes en contextos agrícolas con recursos computacionales limitados. Además, la mayoría de los experimentos carece de esquemas de validación cruzada o protocolos reproducibles, dificultando la replicabilidad de los resultados y su adopción práctica.

Frente a esta brecha, surge la necesidad de realizar evaluaciones sistemáticas que comparen enfoques contemporáneos de aprendizaje profundo bajo un marco experimental controlado. Estos modelos presentan distintos compromisos entre precisión, complejidad y eficiencia computacional, por lo que resulta fundamental analizar su rendimiento en tareas de clasificación automática de plagas y enfermedades agrícolas. Evaluar su comportamiento comparativo permite establecer lineamientos técnicos para su implementación en sistemas

de monitoreo inteligente y diagnóstico fitosanitario automatizado, promoviendo soluciones que combinen alto desempeño, estabilidad y viabilidad operativa en entornos reales de producción.<sup>(31)</sup>

En este estudio se propone un enfoque de visión por computadora impulsado por aprendizaje profundo (Deep Learning-Driven Computer Vision) orientado a la detección temprana y automática de plagas y enfermedades en frutos de cacao. Se implementó un proceso metodológico reproducible que integra estrategias de preprocesamiento, validación cruzada estratificada y análisis estadístico para comparar el rendimiento de diferentes modelos de aprendizaje profundo. El objetivo es identificar patrones visuales asociados a distintas condiciones fitosanitarias y evaluar la estabilidad y precisión de los modelos en escenarios de clasificación automatizada. Este trabajo busca aportar evidencia técnica y práctica para el desarrollo de sistemas inteligentes de diagnóstico agrícola, promoviendo la sostenibilidad y la eficiencia en la producción mundial de cacao.<sup>(32)</sup>

## MÉTODO

El desarrollo experimental se estructuró en un pipeline reproducible de visión por computadora diseñado para la clasificación automática de enfermedades en frutos de cacao. El proceso comprendió las siguientes fases: (1) exploración de datos (EDA), (2) preprocesamiento, (3) entrenamiento con validación cruzada estratificada, (4) evaluación mediante métricas de clasificación y (5) análisis estadístico inferencial para contrastar el desempeño de los modelos. Este esquema metodológico garantiza la trazabilidad de los resultados y la comparabilidad entre configuraciones, permitiendo replicar el estudio en distintos contextos de producción o con nuevos conjuntos de datos.

### Conjunto de datos y preparación inicial

Se utilizó el conjunto de imágenes llamado “Cacao Diseases”, obtenido a través de Kaggle (carpetas por clase), compuesto por 4390 imágenes RGB de frutos de cacao etiquetadas en tres categorías: Black Pod Rot (943), Pod Borer (103) y Healthy (3344). Las imágenes se organizaron bajo el esquema ImageFolder (una carpeta por clase) para su consumo directo por PyTorch. Antes de entrenar, se ejecutó una EDA básica: primero un conteo por clase para identificar desbalances, luego el muestreo de tamaños de imagen para estimar la variabilidad de resoluciones y, por último, un mosaico visual de ejemplos por clase. Cualquier imagen corrupta o ilegible se descartó.

### Diseño experimental y particionamiento

Para obtener estimaciones robustas y habilitar comparación estadística entre arquitecturas, se aplicó validación cruzada estratificada de 5 folds (mismo reparto por clase en cada fold). En cada fold se entrenó y validó cada uno de los tres modelos sobre las mismas particiones, lo que posteriormente permitió tratar al fold como “sujeto” en el análisis de medidas repetidas. Se fijó una semilla global (42) para asegurar reproducibilidad.

### Preprocesamiento y aumentos

Con el fin de preservar señales cromáticas relevantes para síntomas tales como manchas, micelio, o decoloraciones, se usaron parámetros de aumentos moderados y normalización estándar de ImageNet:

Entrenamiento:

```
train_tf = transforms.Compose([
    transforms.Resize(int(img_size*1,15)),
    transforms.RandomResizedCrop(img_size, scale=(0,8, 1,0)),
    transforms.RandomHorizontalFlip(),
    transforms.RandomRotation(10),
    transforms.ToTensor(),
    transforms.Normalize(IMAGENET_MEAN, IMAGENET_STD),
])
```

Validación:

```
eval_tf = transforms.Compose([
    transforms.Resize(int(img_size*1,15)),
    transforms.CenterCrop(img_size),
    transforms.ToTensor(),
    transforms.Normalize(IMAGENET_MEAN, IMAGENET_STD),
])
```

El tamaño de entrada se fijó en 224×224px para permitir la compatibilidad con el estándar de ImageNet y mejorar la eficiencia de cómputo. Las transformaciones en los aumentadores fueron conservadores, con el propósito de no distorsionar rasgos finos en las lesiones.

### Modelos evaluados

Se evaluaron tres arquitecturas de aprendizaje profundo ampliamente utilizadas en estudios recientes de

clasificación de imágenes agro-fitopatológicas: ResNet50<sup>(33)</sup>, EfficientNet-B0<sup>(34)</sup> y Vision Transformer (ViT-B/16).<sup>(35)</sup> Todas las redes fueron preentrenadas en el conjunto de datos ImageNet-1K, con el objetivo de aprovechar las representaciones visuales generales aprendidas a partir de millones de imágenes naturales. Posteriormente, se reconfiguraron las capas de salida para adaptarlas al dominio específico del problema. En el caso de ResNet50, la capa final fue reemplazada por una capa totalmente conectada con tres neuronas correspondientes a las clases de interés. En EfficientNet-B0, se sustituyó la última capa del clasificador por una salida de tres clases, mientras que en ViT-B/16 se reemplazó el componente de clasificación original por una capa lineal con la misma dimensionalidad de salida.

En los tres modelos se realizó un ajuste completo de los parámetros (fine-tuning total), en lugar de congelar las capas convolucionales o de atención preentrenadas. Esta decisión metodológica responde a la diferencia sustancial entre el dominio fuente, representado por ImageNet, basado en objetos naturales y escenas generales, y el dominio objetivo del presente estudio, centrado en imágenes de frutos de cacao afectados por lesiones y plagas. El ajuste integral de pesos permite que las arquitecturas aprendan representaciones discriminativas más específicas del contexto fitopatológico, optimizando la sensibilidad del modelo ante patrones visuales sutiles y texturas asociadas a los diferentes estados de salud del fruto.

### Configuración de entrenamiento

El entrenamiento de los modelos se implementó utilizando el framework PyTorch, aplicando un conjunto uniforme de hiperparámetros con el propósito de garantizar la comparabilidad entre arquitecturas. Se empleó el optimizador AdamW<sup>(36)</sup> con una tasa de aprendizaje inicial de  $1 \times 10^{-4}$  y un parámetro de decaimiento de peso (weight decay) de  $1 \times 10^{-4}$ , configuraciones que equilibran la estabilidad del gradiente y la regularización del modelo durante el aprendizaje. La función de pérdida seleccionada fue la Cross-Entropy Loss, adecuada para problemas de clasificación multiclase. El tamaño de lote se estableció en 32 imágenes por iteración, y el proceso de entrenamiento se ejecutó durante un máximo de 10 épocas por fold, dentro del esquema de validación cruzada estratificada. Para ajustar dinámicamente la tasa de aprendizaje, se incorporó un programador ReduceLROnPlateau, el cual reduce el valor de la tasa de aprendizaje en un 50 % cuando la pérdida de validación no muestra mejora significativa. Asimismo, se implementó un mecanismo de parada temprana con una paciencia de tres épocas, seleccionando como punto óptimo el checkpoint correspondiente al mayor valor de F1 macro alcanzado en la validación. Este conjunto de estrategias permitió controlar el sobreajuste, estabilizar la convergencia y optimizar el uso de recursos computacionales, lo que resulta adecuado tanto para entornos con limitaciones de hardware como para escenarios de inferencia en dispositivos de borde.

### Protocolo de evaluación por fold

Al final de cada época se evaluó en el conjunto de validación del fold y, una vez activado el criterio de parada, se cargó el mejor estado del modelo para computar las métricas finales del fold: Exactitud (Accuracy), F1 macro, Precisión macro y Recall macro. Así como un reporte de clasificación por clase (precision/recall/F1) y la matriz de confusión.

Las matrices de confusión por fold se almacenaron y posteriormente se agregaron por modelo de dos maneras: Por suma de cuentas absolutas y mediante una versión normalizada por fila (recall por clase), con el fin de analizar patrones de confusión sistemáticos entre clases sin que un único fold domine la interpretación.

### Comparación estadística entre modelos

Para contrastar el rendimiento medio de las tres arquitecturas se aplicó un diseño estadístico de medidas repetidas considerando los cinco folds generados en la validación cruzada. En primer lugar, se construyó una tabla en formato largo que incluyó como factores el fold (tratado como sujeto), el modelo (considerado como factor intra-sujeto con tres niveles) y la métrica de respuesta F1 macro. Posteriormente, se realizó un análisis de varianza de medidas repetidas (ANOVA-RM) utilizando el modelo como efecto intra-sujeto y el fold como unidad de observación, estableciendo un nivel de significancia de  $\alpha = 0,05$ . En los casos en que se detectó significancia estadística, se efectuó un análisis post-hoc mediante pruebas t pareadas entre cada par de modelos, aplicando la corrección de Bonferroni para controlar el error tipo I acumulado. Este enfoque permitió evaluar de manera robusta las diferencias atribuibles a la arquitectura, controlando la variabilidad introducida por las particiones de validación y evitando conclusiones dependientes de una única división entrenamiento-validación.

## RESULTADOS

### Exploración de datos (EDA)

Tras la exploración inicial del conjunto de datos se contabilizaron 4 390 imágenes distribuidas en tres clases con un desbalance considerable: la categoría Healthy concentró 3 344 imágenes (76,2 %), mientras que Black Pod Rot presentó 943 (21,5 %) y Pod Borer únicamente 103 (2,35 %), lo que representa una razón aproximada de 32,5 veces entre la clase mayoritaria y la minoritaria. El análisis geométrico realizado sobre

una muestra de 300 imágenes indicó que todas las capturas presentan una relación de aspecto cercana a 1,0, con resoluciones comprendidas entre 1080×1080 y 2160×2160 píxeles, lo que confirma la uniformidad del formato de adquisición. Estas observaciones justificaron la utilización de una validación cruzada estratificada para conservar las proporciones por clase en cada fold, así como la priorización de la métrica F1 macro en la selección de checkpoints con el fin de mitigar el sesgo inducido por el desbalance de clases. Asimismo, se optó por aplicar aumentos de datos moderados durante el entrenamiento, buscando mejorar la capacidad de generalización del modelo sin comprometer la integridad de las señales visuales asociadas a las lesiones (figura 1).

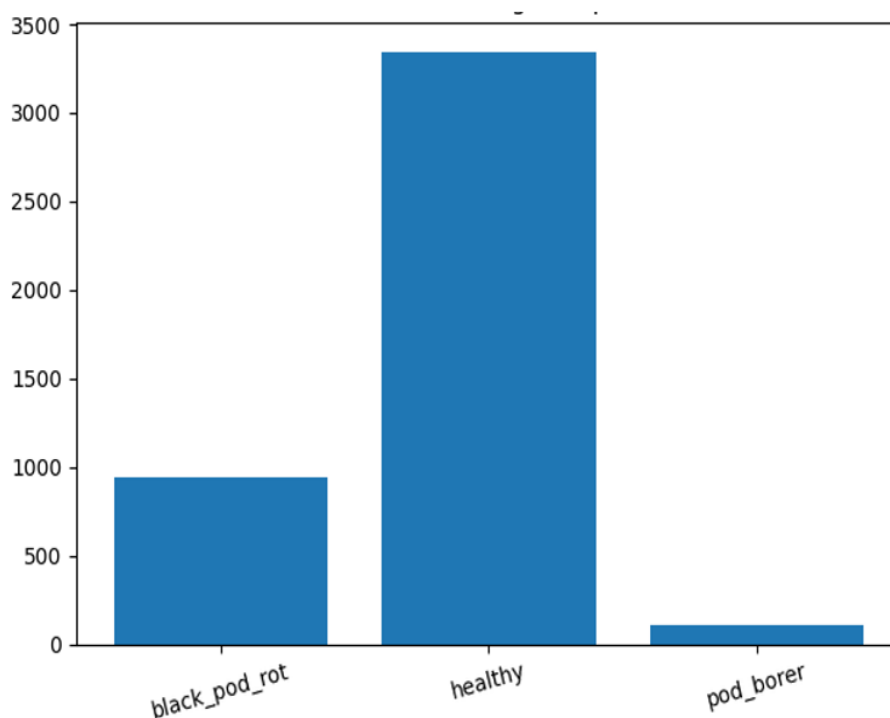


Figura 1. Gráfico de distribución por clases

### Desempeño global de los modelos

El análisis del desempeño global de los modelos evidenció un rendimiento altamente competitivo entre las tres arquitecturas evaluadas. El modelo Vision Transformer (ViT-B/16) alcanzó la mayor media de F1 macro, con un valor de 0,9697 y una desviación estándar de 0,0114, junto con una exactitud promedio de 0,9788. En segundo lugar, se ubicó ResNet50, con una F1 macro media de 0,9683, seguida de EfficientNet-B0, que obtuvo 0,9666. Las diferencias entre los tres modelos resultaron mínimas, con una separación promedio de aproximadamente 0,3 puntos porcentuales entre el mejor y el de menor desempeño en la métrica F1 macro. Este comportamiento estrechamente agrupado refleja la robustez de las arquitecturas y confirma la efectividad del esquema de preprocesamiento y validación adoptado en el estudio (tabla 1).

Tabla 1. Resumen de métricas de evaluación por modelo

| Modelo          | Exac. Media | Desv. Exac. | F1 media | Desv. F1 | Prec. Media | Desv. Prec. | Recall Media | Desv. Recall |
|-----------------|-------------|-------------|----------|----------|-------------|-------------|--------------|--------------|
| EfficientNet-B0 | 0,9756      | 0,0045      | 0,9666   | 0,0128   | 0,9690      | 0,0102      | 0,9651       | 0,0205       |
| ResNet50        | 0,9811      | 0,0036      | 0,9683   | 0,0155   | 0,9803      | 0,0118      | 0,9575       | 0,0233       |
| ViT-B/16        | 0,9788      | 0,0064      | 0,9697   | 0,0114   | 0,9800      | 0,0087      | 0,9604       | 0,0171       |

### Comparación estadística (ANOVA y post-hoc)

La comparación estadística del rendimiento medio entre las tres arquitecturas se realizó mediante un análisis de varianza de medidas repetidas (ANOVA-RM) aplicado sobre la métrica F1 macro, considerando el fold como sujeto y el modelo como factor intra-sujeto. Los resultados no mostraron diferencias estadísticamente significativas entre arquitecturas ( $F = 0,278$ ,  $p = 0,7645$ ), lo que sugiere un comportamiento homogéneo en el rendimiento promedio de los modelos. Este hallazgo se corroboró mediante pruebas post-hoc pareadas con corrección de Bonferroni, en las que ninguno de los contrastes por pares alcanzó significancia estadística. En

particular, la diferencia media de F1 macro entre ViT-B/16 y EfficientNet-B0 fue de aproximadamente +0,0031 con un intervalo de confianza del 95 % y  $p \approx 0,25$ , mientras que la diferencia entre ViT-B/16 y ResNet50 fue de +0,0014, también no significativa. En conjunto, estos resultados indican que, aunque ViT-B/16 presentó la media más alta, las discrepancias observadas fueron pequeñas y estadísticamente atribuibles a la variabilidad del muestreo, confirmando la robustez de las tres arquitecturas evaluadas (tabla 2).

| Tabla 2. Pruebas post-hoc por par de modelos |              |             |
|----------------------------------------------|--------------|-------------|
| Modelos                                      | t            | p_raw       |
| ResNet50 vs EfficientNet B0                  | 0,297661326  | 0,780781239 |
| ResNet50 vs ViT-B/16                         | -0,370845959 | 0,72955701  |
| EfficientNet B0 vs ViT-B/16                  | -1,347132941 | 0,249193532 |

La evolución del F1 macro durante el proceso de entrenamiento y validación mostró un patrón de convergencia estable en los tres modelos, sin indicios de sobreajuste y con fluctuaciones mínimas entre folds. En las figuras 2, 3 y 4 se observa la consistencia de las curvas de validación correspondientes a EfficientNet-B0, ResNet50 y ViT-B/16, respectivamente, lo que confirma la estabilidad del proceso de optimización y la reproducibilidad de los resultados.

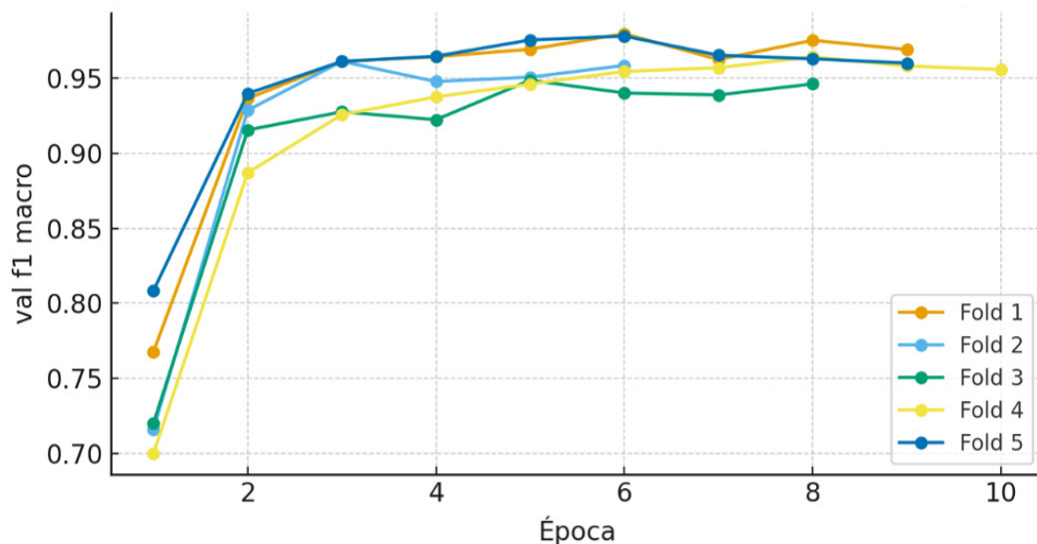


Figura 2. Evolución del F1 Macro (validación) por fold en modelo EfficientNetB0

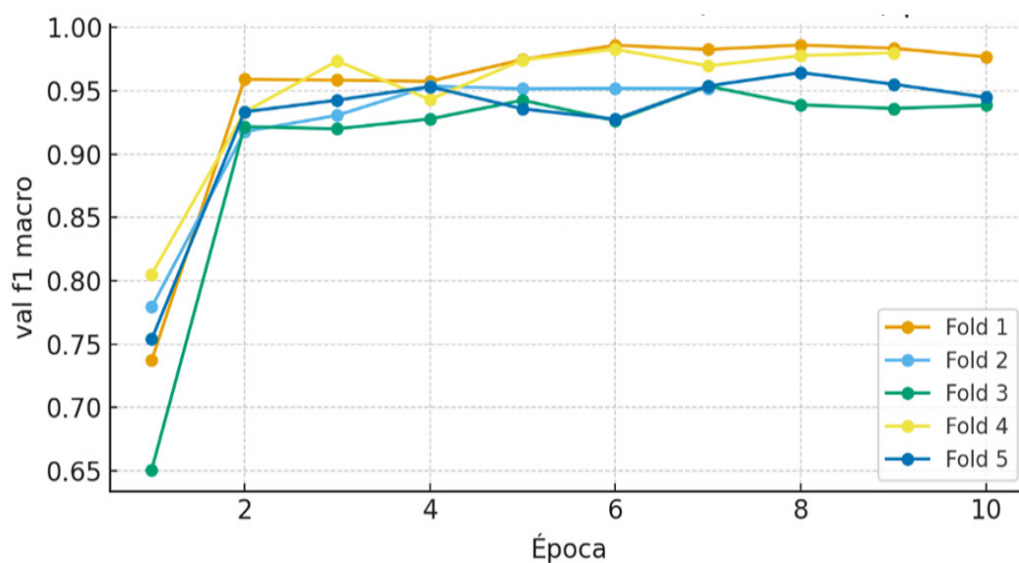


Figura 3. Evolución del F1 Macro (validación) por fold en modelo ResNet50

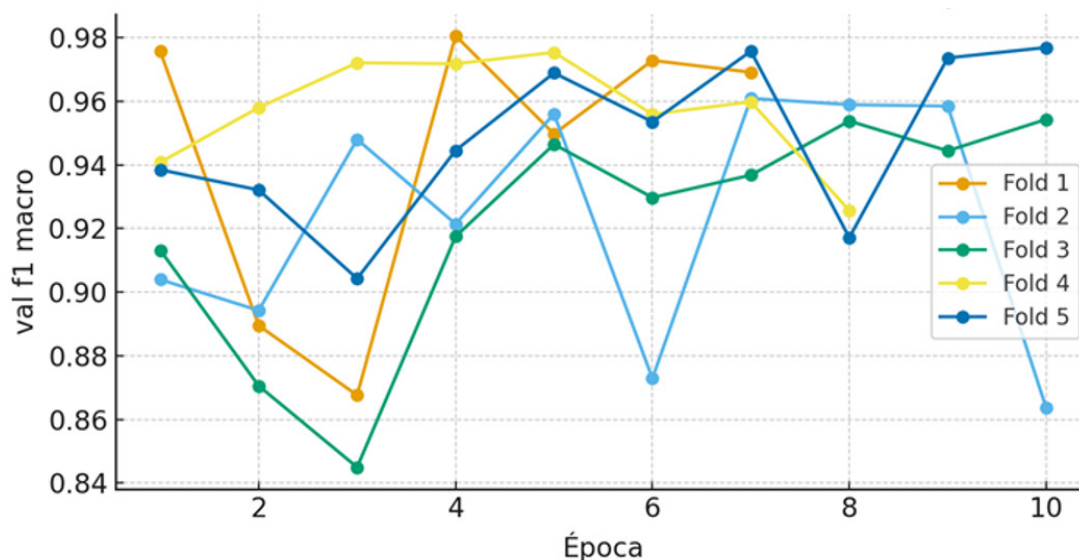


Figura 4. Evolución del F1 Macro (validación) por fold en modelo ViT-B/16

### Consistencia entre folds y dinámica de entrenamiento

El análisis de consistencia entre folds mostró una variabilidad baja en todas las configuraciones experimentales, con desviaciones estándar de la métrica F1 macro cercanas a una centésima, lo que refleja una alta estabilidad del proceso de entrenamiento y validación. Este comportamiento sugiere que las decisiones adoptadas durante el preprocesamiento de las imágenes y el uso de la validación cruzada estratificada aportaron robustez al procedimiento, evitando la influencia de un fold dominante que pudiera sesgar los promedios globales. Los resultados detallados de cada arquitectura en los cinco folds se presentan en la tabla 3, donde se observa la consistencia de las métricas de exactitud, precisión, recall y F1, evidenciando un rendimiento equilibrado entre modelos y particiones.

Asimismo, las curvas de entrenamiento y validación mostraron una convergencia rápida, alcanzando estabilidad en menos de diez épocas. El uso del programador de tasa de aprendizaje (scheduler) favoreció la estabilización de la pérdida en las etapas finales del entrenamiento, mientras que el criterio de parada temprana contribuyó a prevenir el sobreajuste. En conjunto, estos resultados confirman la reproducibilidad del pipeline propuesto y la consistencia inter-fold de los modelos, lo que respalda la validez de las conclusiones derivadas del análisis comparativo.

Tabla 3. Métricas de validación cruzada estratificada

| Fold | Modelo          | Exactitud | F1     | Precisión | Recall |
|------|-----------------|-----------|--------|-----------|--------|
| 1    | ResNet50        | 0,9863    | 0,9862 | 0,9914    | 0,9814 |
| 1    | EfficientNet-B0 | 0,9795    | 0,9799 | 0,9765    | 0,9834 |
| 1    | ViT-B/16        | 0,9806    | 0,9807 | 0,9825    | 0,9789 |
| 2    | ResNet50        | 0,9795    | 0,9539 | 0,9820    | 0,9299 |
| 2    | EfficientNet-B0 | 0,9692    | 0,9614 | 0,9557    | 0,9676 |
| 2    | ViT-B/16        | 0,9681    | 0,9610 | 0,9665    | 0,9560 |
| 3    | ResNet50        | 0,9772    | 0,9538 | 0,9601    | 0,9479 |
| 3    | EfficientNet-B0 | 0,9727    | 0,9488 | 0,9709    | 0,9305 |
| 3    | ViT-B/16        | 0,9784    | 0,9543 | 0,9781    | 0,9343 |
| 4    | ResNet50        | 0,9829    | 0,9829 | 0,9835    | 0,9823 |
| 4    | EfficientNet-B0 | 0,9784    | 0,9645 | 0,9616    | 0,9675 |
| 4    | ViT-B/16        | 0,9829    | 0,9754 | 0,9899    | 0,9619 |
| 5    | ResNet50        | 0,9795    | 0,9645 | 0,9844    | 0,9463 |
| 5    | EfficientNet-B0 | 0,9784    | 0,9783 | 0,9801    | 0,9765 |
| 5    | ViT-B/16        | 0,9841    | 0,9769 | 0,9830    | 0,9713 |

Las matrices de confusión agregadas por modelo, obtenidas a partir de la suma de los cinco folds, junto

con sus versiones normalizadas por fila, evidenciaron un comportamiento equilibrado entre clases, sin indicios de patrones de confusión sistemáticos dominantes. La normalización por fila permitió analizar de manera detallada el recall por clase, mostrando ligeras oscilaciones entre arquitecturas que resultan coherentes con el rango estrecho de valores de F1 macro previamente reportado. Estas variaciones no modifican la conclusión general del estudio: los tres modelos presentan una capacidad consistente para discriminar correctamente entre las categorías Black Pod Rot, Pod Borer y Healthy, manteniendo un desempeño homogéneo en el conjunto de datos evaluado (figura 5).

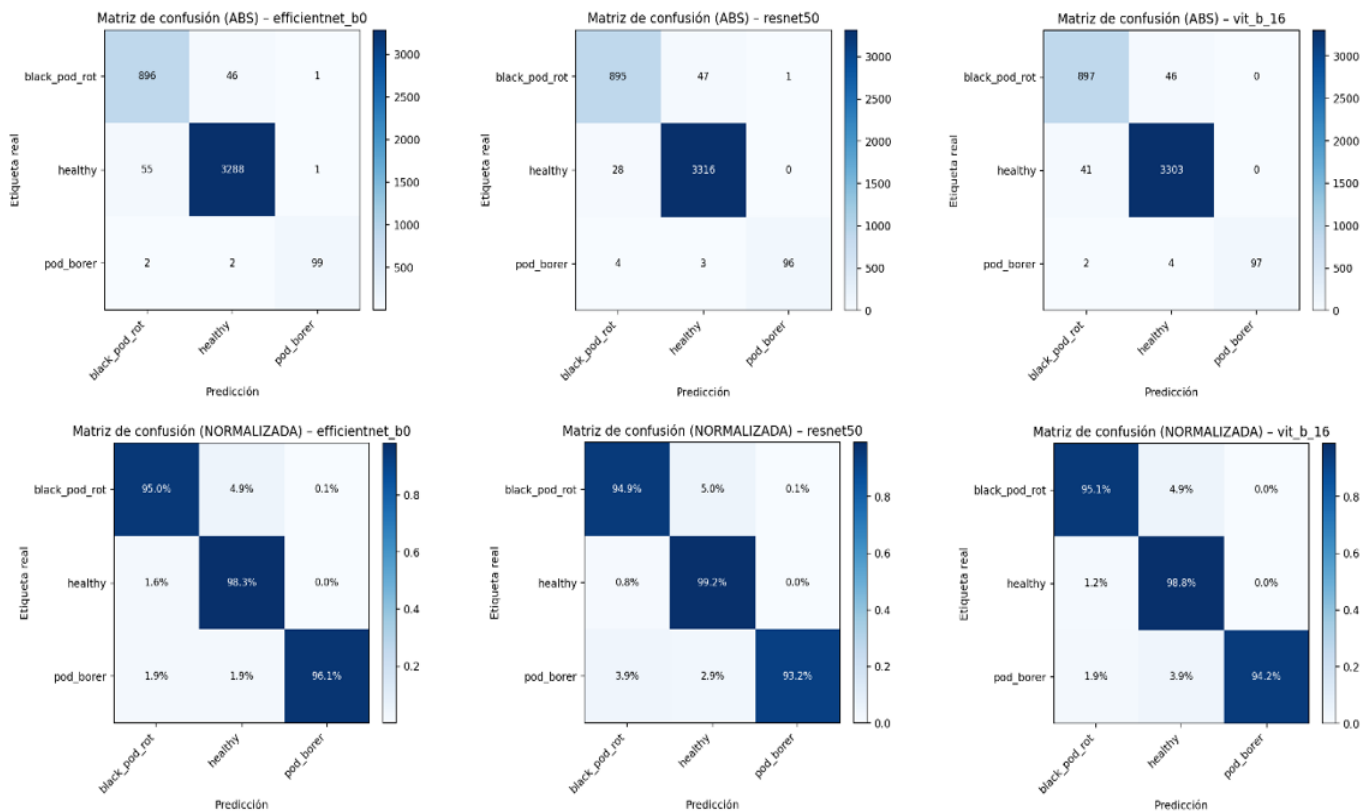


Figura 5. Matrices de confusión por modelo

## DISCUSIÓN

Los resultados demuestran que las tres arquitecturas evaluadas ofrecen un desempeño competitivo y estable para la clasificación automática de enfermedades del cacao, con valores de F1 macro cercanos y sin diferencias estadísticamente significativas. Este hallazgo confirma que la eficacia del modelo no depende estrictamente de la arquitectura empleada, sino de la coherencia del pipeline de preprocesamiento, la validación estratificada y el control de sobreajuste. En consecuencia, la elección del modelo de referencia puede guiarse por criterios operativos y contextuales de implementación: Vision Transformer (ViT-B/16), por su ligera ventaja media y capacidad para modelar relaciones espaciales globales; ResNet50, por su madurez, amplio soporte en bibliotecas de producción y comportamiento muy cercano al primero; y EfficientNet-B0, cuando se prioriza la eficiencia computacional y la reducción de la huella de despliegue.

Estos resultados respaldan lo señalado por Ray et al.<sup>(25)</sup> y Deepa et al.<sup>(28)</sup>, quienes destacan que los modelos de aprendizaje profundo aplicados al diagnóstico agrícola ofrecen altos niveles de precisión incluso con arquitecturas ligeras, siempre que se acompañen de estrategias de entrenamiento consistentes y datos adecuadamente preprocesados. Asimismo, confirman la tendencia observada por Lebrini y Ayerdi Gotor<sup>(21)</sup> y Shafay et al.<sup>(29)</sup>, respecto a que la eficacia de la visión por computadora en el ámbito fitosanitario depende más del rigor metodológico que de la complejidad del modelo en sí.

Desde el punto de vista aplicado, los resultados contribuyen a llenar un vacío metodológico identificado en la literatura reciente: la falta de comparaciones sistemáticas entre arquitecturas contemporáneas bajo marcos estadísticos controlados. En este sentido, el estudio aporta evidencia reproducible que orienta la selección de modelos según el equilibrio entre precisión, estabilidad y viabilidad de implementación en entornos agrícolas de recursos limitados, una línea de investigación destacada por Bono et al.<sup>(31)</sup> en el contexto de la agricultura inteligente. Finalmente, las mejoras futuras deberían enfocarse no tanto en el reemplazo de arquitecturas, sino en optimizar estrategias complementarias como el rebalanceo de clases, los aumentos de datos específicos por tipo de lesión y el ajuste adaptativo de umbrales de decisión, siguiendo las recomendaciones de Song et

al.<sup>(22)</sup> sobre la necesidad de integrar pipelines robustos que maximicen la generalización en aplicaciones de visión por computadora agrícola.

## CONCLUSIONES

El estudio demostró que las arquitecturas de aprendizaje profundo evaluadas presentan un rendimiento estadísticamente equivalente en la detección automática de plagas y enfermedades del cacao, alcanzando valores de F1 macro superiores al 0,96 y desviaciones estándar reducidas entre folds. Estos resultados evidencian la robustez del pipeline de visión por computadora implementado y la eficacia de las estrategias de preprocesamiento, validación y regularización empleadas. La ausencia de diferencias significativas entre modelos indica que la selección del enfoque puede guiarse por criterios operativos, considerando el balance entre precisión, eficiencia computacional y escalabilidad del despliegue. Asimismo, se confirma que la estabilidad y generalización del sistema dependen principalmente de la calidad del entrenamiento y del manejo del desbalance de clases, más que del tipo de arquitectura utilizada. En conjunto, los hallazgos contribuyen al desarrollo de sistemas inteligentes reproducibles y escalables para el monitoreo fitosanitario, fortaleciendo la integración de la inteligencia artificial en la agricultura de precisión y su aplicación en contextos productivos reales.

## REFERENCIAS

1. Charry A, Perea C, Ramírez K, Zambrano G, Yovera F, Santos A, et al. The bittersweet economics of different cacao production systems in Colombia, Ecuador and Peru. *Agric Syst.* 2025;224:104235. <https://doi.org/10.1016/j.agsy.2024.104235>.
2. Vinci G, Ruggeri M, Gobbi L, Savastano M. Social Life Cycle Assessment of Cocoa Production: Evidence from Ivory Coast and Ghana. *Resources.* 2024;13:141. <https://doi.org/10.3390/resources13100141>.
3. Angulo Rincón SO, Solarte Solarte CM. Green innovation as a sustainable alternative to strengthen cocoa cultivation. *Environ Res Ecotoxicity.* 2026;5:300. <https://doi.org/10.56294/ere2026300>.
4. Angulo Rincón SO, Solarte Solarte CM. Green innovation analysis for cocoa farmers in the municipality of Roberto Payán. *Environ Res Ecotoxicity.* 2024;3:120. <https://doi.org/10.56294/ere2024120>.
5. Quintero I, Ceccaldi A, Martínez H, Santander M, Rodríguez J, Escobar S. Dry cacao pulp in chocolate bars: A sustainable, nutrient-rich sweetener with enhanced sensory quality through refractance windows drying. *Appl Food Res.* 2025;5:100700. <https://doi.org/10.1016/j.afres.2025.100700>.
6. Paparella A, Schirone M, López CC. The Health Impact of Cocoa from Cultivation to the Formation of Biogenic Amines: An Updated Review. *Foods.* 2025;14:255. <https://doi.org/10.3390/foods14020255>.
7. Delgado-Ospina J, Molina-Hernández JB, Chaves-López C, Romanazzi G, Paparella A. The Role of Fungi in the Cocoa Production Chain and the Challenge of Climate Change. *J Fungi.* 2021;7:202. <https://doi.org/10.3390/jof7030202>.
8. Zahlul Ikhsan, Muhammad Al-Ikhlās, Yaherwandi, Hasmiandy Hamid, Aulia Oktavia, Dandy Ahamefula Osibe. Level of Pest Infestation on Cocoa (*Theobroma cacao* L.) Variety BL-50 in Tanah Datar Regency, West Sumatera Province, Indonesia. *Andalasian Int J Entomol.* 2024;2:38-47. <https://doi.org/10.25077/aijent.2.1.38-47.2024>.
9. Cilas C, Bastide P. Challenges to Cocoa Production in the Face of Climate Change and the Spread of Pests and Diseases. *Agronomy.* 2020;10:1232. <https://doi.org/10.3390/agronomy10091232>.
10. Puig AS, Irish B, Ayala-Silva T, Wurzel S, Gutierrez O. Effect of Cacao Black Pod Rot Screening Method on Disease Reaction Determination. 1st Int Online Conf Agric Agric Sci Technol. Basel Switzerland: MDPI; 2022. p. 71. <https://www.mdpi.com/2673-4583/10/1/71>.
11. Schmidt JE, Puig AS, DuVal AE, Pfeufer EE. Phyllosphere microbial diversity and specific taxa mediate within-cultivar resistance to *Phytophthora palmivora* in cacao. *mSphere.* 2023;8. <https://doi.org/10.1128/msphere.00013-23>.
12. Magfirah N, Sartiami D, Niogret J, Ekayanti A. Biological parameter of cocoa pod borer *Conopomorpha cramerella* (Lepidoptera: Gracillariidae) under laboratory conditions. *IOP Conf Ser Earth Environ Sci.*

2025;1494:012020. <https://doi.org/10.1088/1755-1315/1494/1/012020>.

13. Miyittah MK, Kosivi RK, Tulashie SK, Addi MN, Tawiah JY. The need for alternative pest management methods to mitigate risks among cocoa farmers in the Volta region, Ghana. *Heliyon*. 2022;8:e12591. <https://doi.org/10.1016/j.heliyon.2022.e12591>.
14. Polania Bello V. Diagnosis of the agronomic and phytosanitary management practices of cocoa producers in San José del Fragua: the case of the Cerafín García property. *Environ Res Ecotoxicity*. 2023;2:55. <https://doi.org/10.56294/ere202355>.
15. Wu K, Ji Z, Wang H, Shao X, Li H, Zhang W, et al. A Comprehensive Review of AI Methods in Agri-Food Engineering: Applications, Challenges, and Future Directions. *Electronics*. 2025;14:3994. <https://doi.org/10.3390/electronics14203994>.
16. Raj M, Prahadeeswaran M. Revolutionizing agriculture: a review of smart farming technologies for a sustainable future. *Discov Appl Sci*. 2025;7:937. <https://doi.org/10.1007/s42452-025-07561-6>.
17. Tirado J. Big data and artificial intelligence for innovation in management in the agricultural industry in Extremadura. *EthAlca*. 2022;1:38. <https://doi.org/10.56294/ai202238>.
18. Waqas M, Naseem A, Humphries UW, Hlaing PT, Dechpichai P, Wangwongchai A. Applications of machine learning and deep learning in agriculture: A comprehensive review. *Green Technol Sustain*. 2025;3:100199. <https://doi.org/10.1016/j.grets.2025.100199>.
19. Taha MF, Mao H, Zhang Z, Elmasry G, Awad MA, Abdalla A, et al. Emerging Technologies for Precision Crop Management Towards Agriculture 5.0: A Comprehensive Overview. *Agriculture*. 2025;15:582. <https://doi.org/10.3390/agriculture15060582>.
20. Injante R, Rios-Trigoso G, Ramírez-Shupingahua S, Tejada Shupingahua K. Procesamiento de imágenes para la detección de la madurez del fruto: una revisión sistemática. *Rev Peru Ing Arquít y Medio Ambient*. 2025;2. <https://doi.org/10.37711/repiama.2025.2.1.3>.
21. Lebrini Y, Ayerdi Gotor A. Crops Disease Detection, from Leaves to Field: What We Can Expect from Artificial Intelligence. *Agronomy*. 2024;14:2719. <https://doi.org/10.3390/agronomy14112719>.
22. Song X, Yan L, Liu S, Gao T, Han L, Jiang X, et al. Agricultural Image Processing: Challenges, Advances, and Future Trends. *Appl Sci*. 2025;15:9206. <https://doi.org/10.3390/app15169206>.
23. Anjali, Jena A, Bamola A, Mishra S, Jain I, Pathak N, et al. State-of-the-art non-destructive approaches for maturity index determination in fruits and vegetables: principles, applications, and future directions. *Food Prod Process Nutr*. 2024;6:56. <https://doi.org/10.1186/s43014-023-00205-5>.
24. Uriarte del Águila C. Application of data mining as a strategic tool in knowledge management. *Diginomics*. 2024;3:123. <https://doi.org/10.56294/digi2024123>.
25. Ray RK, Chakravarty S, Dash S, Ghosh A, Mohanty SN, Reddy Chirra VR, et al. Precision pest management in agriculture using Inception V3 and EfficientNet B4: A deep learning approach for crop protection. *Inf Process Agric*. 2025. <https://doi.org/10.1016/j.inpa.2025.09.005>.
26. Villalobos-Culqui C, Valles-Coral MA, García-Rivas-Plata C, Tuesta-Hidalgo OA. Modelo de visión artificial basada en redes neuronales convolucionales para identificación de mazorca negra en plantaciones de cacao. *Rev Científica Sist e Informática*. 2025;5:e678. <https://doi.org/10.51252/rcsi.v5i1.678>.
27. Mall PK, Singh PK, Srivastav S, Narayan V, Paprzycki M, Jaworska T, et al. A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthc Anal*. 2023;4:100216. <https://doi.org/10.1016/j.health.2023.100216>.
28. Deepa R, Varun P, Shoba LK, Swathy R, Prabhu B. Advanced Image Processing Techniques for Precision Agriculture. *Innov Dev Unmanned Aer Veh*. IGI Global Scientific Publishing; 2025. p. 227-244. <https://services.igi-global.com/viewfullarticle.aspx?doi=10.4018/9781668414444.ch014>.

igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/979-8-3693-8462-6.ch010.

29. Shafay M, Hassan T, Owais M, Hussain I, Khawaja SG, Seneviratne L, et al. Recent advances in plant disease detection: challenges and opportunities. *Plant Methods*. 2025;21:140. <https://doi.org/10.1186/s13007-025-01450-0>.

30. Jiménez Zapata EM. Rural socioeconomic transformations mediated by AI. *EthAlca*. 2025;4:156. <https://doi.org/10.56294/ai2025156>.

31. Bono A, Guaragnella C, D'Orazio T. A perspective analysis of imaging-based monitoring systems in precision viticulture: Technologies, intelligent data analyses and research challenges. *Artif Intell Agric*. 2026;16:62-84. <https://doi.org/10.1016/j.aiia.2025.08.001>.

32. Tirado J. Agriculture in the digital age: organizational challenges for the adoption of AI and Big Data. *EthAlca*. 2022;1:17. <https://doi.org/10.56294/ai202217>.

33. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proc IEEE Conf Comput Vis pattern Recognit*. 2015. p. 770-778. <http://arxiv.org/abs/1512.03385>.

34. Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Int Conf Mach Learn*. 2020. p. 6105-6114. <http://arxiv.org/abs/1905.11946>.

35. Wang Y, Deng Y, Zheng Y, Chattopadhyay P, Wang L. Vision Transformers for Image Classification: A Comparative Survey. *Technologies*. 2025;13:32. <https://doi.org/10.3390/technologies13010032>.

36. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *3rd Int Conf Learn Represent*. 2014. <http://arxiv.org/abs/1412.6980>.

## FINANCIACIÓN

Los autores no recibieron financiación para el desarrollo de la presente investigación.

## CONFLICTO DE INTERESES

Los autores declaran que no existe conflicto de intereses.

## CONTRIBUCIÓN DE AUTORÍA

*Conceptualización:* Jorge Raúl Navarro-Cabrera.

*Curación de datos:* Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes.

*Análisis formal:* Jorge Raúl Navarro-Cabrera, Ángel Cárdenas-García.

*Investigación:* Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes, Ángel Cárdenas-García, Carlos Mauricio Lozano-Carranza.

*Metodología:* Jorge Raúl Navarro-Cabrera.

*Software:* Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes.

*Validación:* Jorge Raúl Navarro-Cabrera, Carlos Mauricio Lozano-Carranza.

*Visualización:* Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes, Ángel Cárdenas-García.

*Redacción - borrador original:* Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes, Carlos Mauricio Lozano-Carranza.

*Redacción - revisión y edición:* Jorge Raúl Navarro-Cabrera, Ángel Cárdenas-García.