



# Aplicación de aprendizaje profundo en la detección fitosanitaria del cacao usando visión por computadora

Application of deep learning in the phytosanitary detection of cocoa using computer vision

Navarro-Cabrera, Jorge Raul<sup>1\*</sup>

Beraún-Barrantes, José Guillermo<sup>1</sup>

Cárdenas-García, Ángel<sup>2</sup>

Lozano-Carranza, Carlos Mauricio<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

**Recibido:** 16 Set. 2025 | **Aceptado:** 30 Dic. 2025 | **Publicado:** 20 Ene. 2026

**Autor de correspondencia\*:** [jorge.navarro@udh.edu.pe](mailto:jorge.navarro@udh.edu.pe)

**Cómo citar este artículo:** Navarro-Cabrera, J. R., Beraún-Barrantes, J. G., Cárdenas-García, A. & Lozano-Carranza, C. M. (2026).

Aplicación de aprendizaje profundo en la detección fitosanitaria del cacao usando visión por computadora. *Revista Científica de Sistemas e Informática*, 6(1), e1385. <https://doi.org/10.51252/rcsi.v6i1.1385>

## RESUMEN

Este estudio analizó la aplicación del aprendizaje profundo en la detección fitosanitaria automatizada del cacao mediante visión por computadora, comparando el desempeño de tres arquitecturas: ResNet50, EfficientNet-B0 y Vision Transformer (ViT-B/16). Se implementó un pipeline reproducible que integró preprocesamiento de imágenes, validación cruzada estratificada de cinco pliegues y análisis estadístico inferencial mediante ANOVA de medidas repetidas. El conjunto de datos estuvo conformado por 4390 imágenes RGB de frutos de cacao, distribuidas en tres clases desbalanceadas: Healthy, Black Pod Rot y Pod Borer. Todos los modelos fueron ajustados mediante fine-tuning completo y entrenados con el optimizador AdamW, parada temprana y programación dinámica de la tasa de aprendizaje. Los resultados mostraron valores medios de F1 macro superiores a 0.96 en las tres arquitecturas, sin diferencias estadísticamente significativas entre modelos ( $F = 0,278$ ,  $p = 0,7645$ ). Las curvas de entrenamiento evidenciaron convergencia estable y baja variabilidad inter-fold, sin indicios de sobreajuste. Los hallazgos indican que el rendimiento depende principalmente de la calidad del pipeline experimental y del manejo del desbalance de clases, más que del tipo de arquitectura empleada.

**Palabras clave:** aprendizaje profundo; diagnóstico automatizado; modelos de clasificación; sanidad vegetal; visión artificial

## ABSTRACT

This study analyzed the application of deep learning for automated phytosanitary detection in cacao using computer vision, comparing the performance of three architectures: ResNet50, EfficientNet-B0, and Vision Transformer (ViT-B/16). A reproducible pipeline was implemented, integrating image preprocessing, five-fold stratified cross-validation, and inferential statistical analysis using repeated-measures ANOVA. The dataset consisted of 4390 RGB images of cacao fruits distributed across three imbalanced classes: Healthy, Black Pod Rot, and Pod Borer. All models were fully fine-tuned and trained using the AdamW optimizer, early stopping, and a dynamic learning rate scheduler. The results showed mean F1-macro values above 0.96 across all architectures, with no statistically significant differences among models ( $F = 0.278$ ,  $p = 0.7645$ ). Training curves exhibited stable convergence and low inter-fold variability, with no evidence of overfitting. These findings indicate that system performance primarily depends on the quality of the experimental pipeline and class imbalance handling rather than on the specific architecture employed.

**Keywords:** deep learning; automated diagnosis; classification models; plant health; artificial vision



## 1. 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 (Charry et al., 2025; Vinci et al., 2024). Su grano fermentado y seco constituye la materia prima esencial para la industria del chocolate y diversos productos de confitería (Paparella et al., 2025; Quintero et al., 2025). 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 (Cilas & Bastide, 2020; Delgado-Ospina et al., 2021; Zahlul Ikhsan et al., 2024). 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 (Magfirah et al., 2025; Puig et al., 2022; Schmidt et al., 2023).

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 (Miyittah et al., 2022; Polania Bello, 2023). 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 (Raj & Prahadeeswaran, 2025; Wu et al., 2025).

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 (Taha et al., 2025; Waqas et al., 2025). Estas tecnologías permiten analizar grandes volúmenes de datos visuales mediante algoritmos capaces de reconocer patrones complejos en imágenes de hojas, frutos o tallos (Injante et al., 2025; Lebrini & Ayerdi Gotor, 2024). 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 (Anjali et al., 2024; Song et al., 2025).

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 (Ray et al., 2025; Villalobos-Culqui et al., 2025). 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 (Mall et al., 2023). 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 (Deepa et al., 2025; Shafay et al., 2025).

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 (Bono et al., 2026).

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.

## 2. MATERIALES Y MÉTODOS

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. En la figura 1 se muestra el 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.

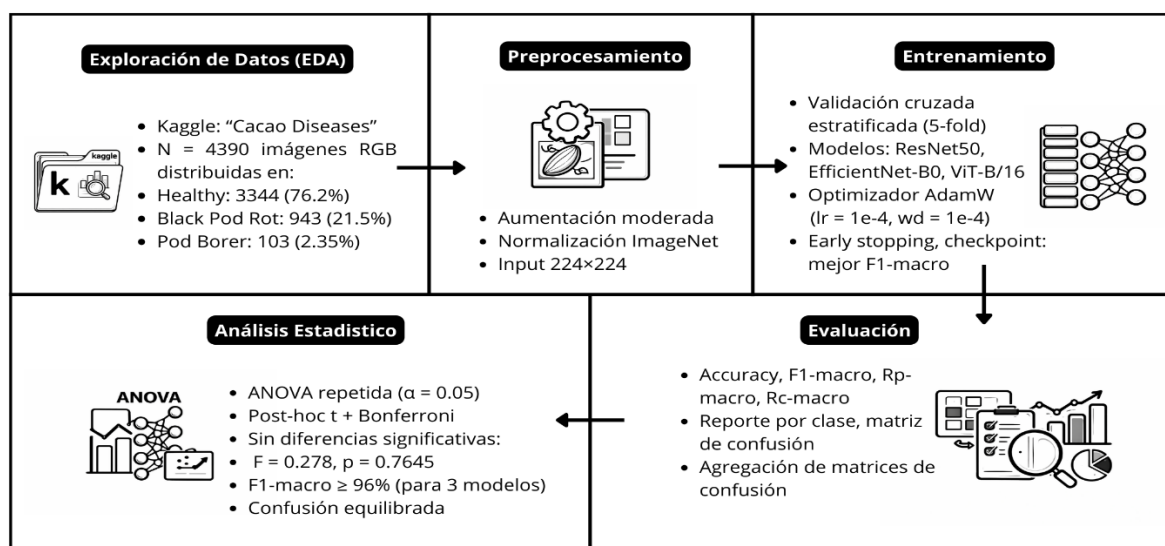


Figura 1. Esquema metodológico propuesto

### 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 (He et al., 2015), EfficientNet-B0 (Tan & Le, 2020) y Vision Transformer (ViT-B/16) (Wang et al., 2025). 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 (Kingma & Ba, 2014) 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

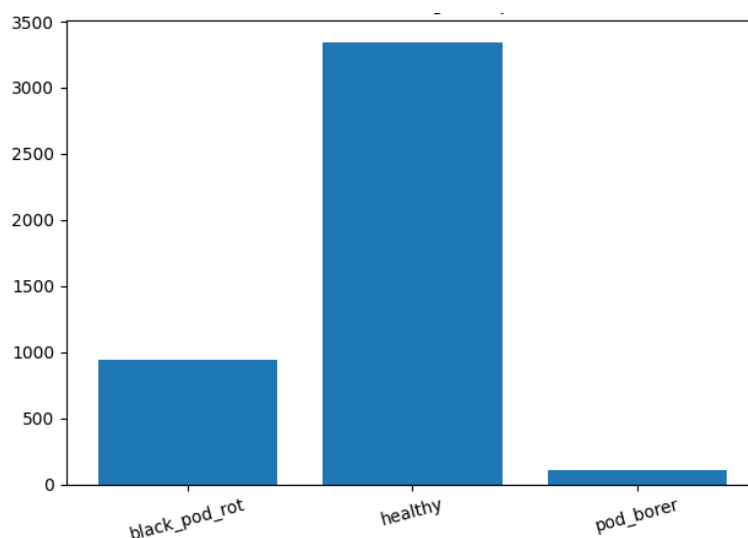
Para contrastar el rendimiento medio de las tres arquitecturas se aplicó un diseño estadístico de medidas repetidas considerando los cinco folds generados en la validación cruzada. En primer lugar, se construyó una tabla en formato largo que incluyó como factores el fold (tratado como sujeto), el modelo (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.

## 3. RESULTADOS Y DISCUSIÓN

### 3.1. 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 2).



**Figura 2.** Gráfico de distribución por clases

### 3.2. 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).

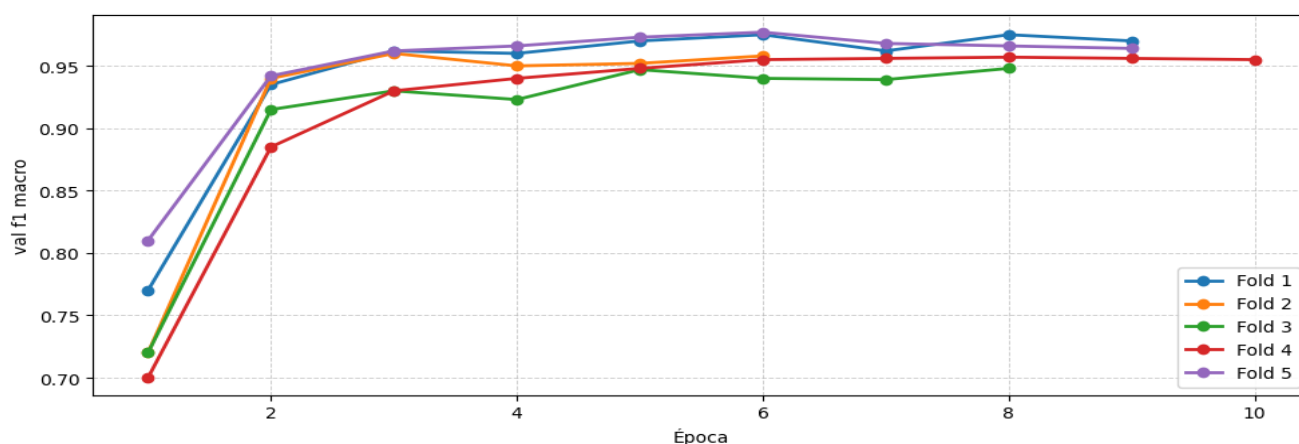
**Tabla 1.** Resumen de métricas de evaluación por modelo

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

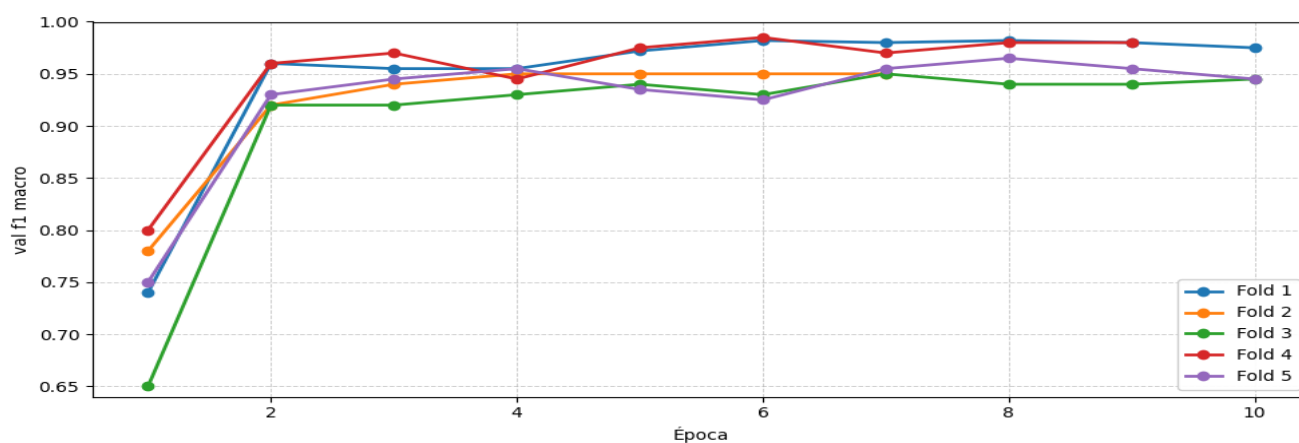
### 3.3. Entrenamiento, validación cruzada estratificada y consistencia inter-fold

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 3, 4 y 5 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 obtenidos.

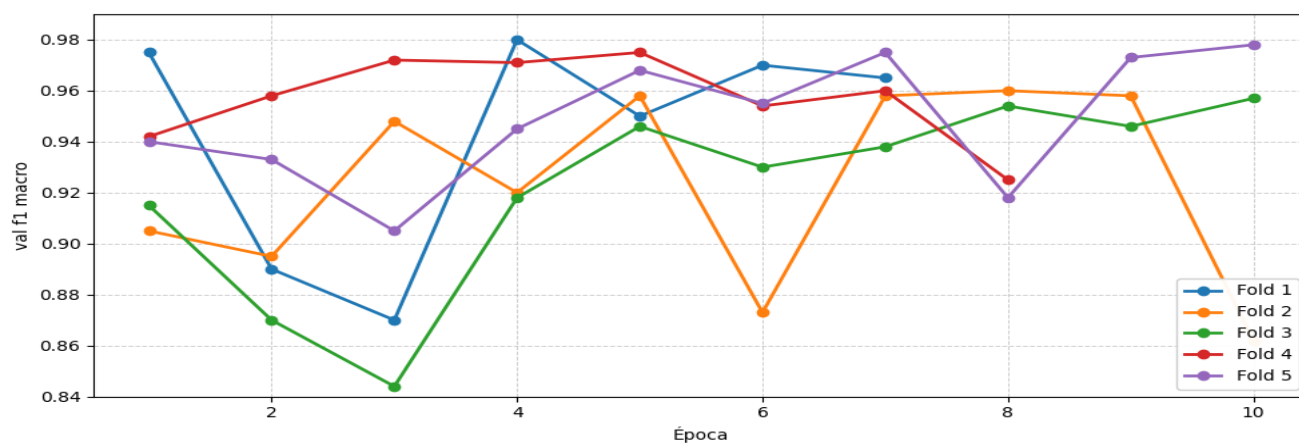




**Figura 3.** Evolución del F1 Macro (validación) por fold en modelo EfficientNetB0



**Figura 4.** Evolución del F1 Macro (validación) por fold en modelo ResNet50



**Figura 5.** Evolución del F1 Macro (validación) por fold en modelo ViT-B/16

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 2, 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 2. 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

3.4. Matrices de confusión y comportamiento por clase

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 6).

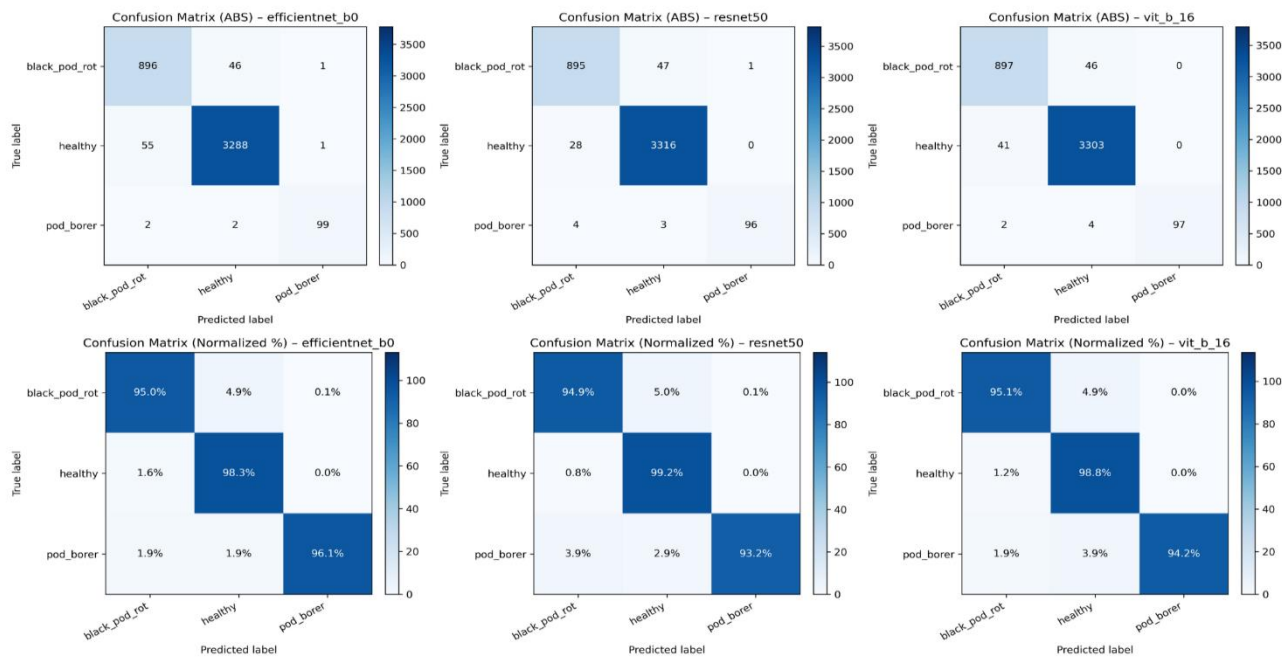


Figura 6. Matrices de confusión por modelo

3.5. Análisis estadístico inferencial (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 3).

**Tabla 3.** 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

### Implicaciones prácticas

Los resultados obtenidos 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. (2025) y Deepa et al. (2025), 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 & Ayerdi Gotor (2024) y Shafay et al. (2025), 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. (2026) 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. (2025) 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.

## FINANCIAMIENTO

Los autores no recibieron ningún patrocinio para llevar a cabo este estudio-artículo.

## CONFLICTO DE INTERESES

Los autores declaran que no existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

## CONTRIBUCIÓN DE AUTORES

Conceptualización: Navarro-Cabrera, J. R. Curación de datos y análisis formal: Beraún-Barrantes, J. G.; Cárdenas-García, Á.; Lozano-Carranza, C. M. Investigación: Navarro-Cabrera, J. R.; Beraún-Barrantes, J. G. Metodología y administración del proyecto: Navarro-Cabrera, J. R. Software: Navarro-Cabrera, J. R. Validación: Navarro-Cabrera, J. R.; Beraún-Barrantes, J. G. Visualización: Navarro-Cabrera, J. R. Redacción – borrador original: Navarro-Cabrera, J. R.; Beraún-Barrantes, J. G.; Cárdenas-García, Á.; Lozano-Carranza, C. M. Redacción – revisión y edición: Navarro-Cabrera, J. R.

## REFERENCIAS

- Anjali, Jena, A., Bamola, A., Mishra, S., Jain, I., Pathak, N., Sharma, N., Joshi, N., Pandey, R., Kaparwal, S., Yadav, V., Gupta, A. K., Jha, A. K., Bhatt, S., Kumar, V., Naik, B., Rustagi, S., Preet, M. S., & Akhtar, S. (2024). State-of-the-art non-destructive approaches for maturity index determination in fruits and vegetables: principles, applications, and future directions. *Food Production, Processing and Nutrition*, 6(1), 56. <https://doi.org/10.1186/s43014-023-00205-5>
- Bono, A., Guaragnella, C., & D'Orazio, T. (2026). A perspective analysis of imaging-based monitoring systems in precision viticulture: Technologies, intelligent data analyses and research challenges. *Artificial Intelligence in Agriculture*, 16(1), 62–84. <https://doi.org/10.1016/j.aiia.2025.08.001>
- Charry, A., Perea, C., Ramírez, K., Zambrano, G., Yovera, F., Santos, A., Jiménez, T., Romero, M., Lundy, M., Quintero, M., & Pulleman, M. (2025). The bittersweet economics of different cacao production systems in Colombia, Ecuador and Peru. *Agricultural Systems*, 224, 104235. <https://doi.org/10.1016/j.agsy.2024.104235>
- Cilas, C., & Bastide, P. (2020). Challenges to Cocoa Production in the Face of Climate Change and the Spread of Pests and Diseases. *Agronomy*, 10(9), 1232. <https://doi.org/10.3390/agronomy10091232>
- Deepa, R., Varun, P., Shoba, L. K., Swathy, R., & Prabhu, B. (2025). Advanced Image Processing Techniques

- for Precision Agriculture. In *Innovations and Developments in Unmanned Aerial Vehicles* (pp. 227–244). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-8462-6.ch010>
- Delgado-Ospina, J., Molina-Hernández, J. B., Chaves-López, C., Romanazzi, G., & Paparella, A. (2021). The Role of Fungi in the Cocoa Production Chain and the Challenge of Climate Change. *Journal of Fungi*, 7(3), 202. <https://doi.org/10.3390/jof7030202>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. In *Cornell University*. <https://doi.org/https://doi.org/1512.03385>
- Injante, R., Rios-Trigoso, G., Ramírez-Shupingahua, S., & Tejada Shupingahua, K. (2025). Procesamiento de imágenes para la detección de la madurez del fruto: una revisión sistemática. *Revista Peruana de Ingeniería, Arquitectura y Medio Ambiente*, 2(1). <https://doi.org/10.37711/repiana.2025.2.1.3>
- Kingma, J., & Ba, D. P. (2014). *Adam: A Method for Stochastic Optimization*. Cornell University. <https://doi.org/https://doi.org/10.48550/arXiv.1412.6980>
- Lebrini, Y., & Ayerdi Gotor, A. (2024). Crops Disease Detection, from Leaves to Field: What We Can Expect from Artificial Intelligence. *Agronomy*, 14(11), 2719. <https://doi.org/10.3390/agronomy14112719>
- Magfirah, N., Sartiami, D., Niogret, J., & Ekayanti, A. (2025). Biological parameter of cocoa pod borer *Conopomorpha cramerella* (Lepidoptera: Gracillariidae) under laboratory conditions. *IOP Conference Series: Earth and Environmental Science*, 1494(1), 012020. <https://doi.org/10.1088/1755-1315/1494/1/012020>
- Mall, P. K., Singh, P. K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T., & Ganzha, M. (2023). A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthcare Analytics*, 4, 100216. <https://doi.org/10.1016/j.health.2023.100216>
- Miyittah, M. K., Kosivi, R. K., Tulashie, S. K., Addi, M. N., & Tawiah, J. Y. (2022). The need for alternative pest management methods to mitigate risks among cocoa farmers in the Volta region, Ghana. *Heliyon*, 8(12), e12591. <https://doi.org/10.1016/j.heliyon.2022.e12591>
- Paparella, A., Schirone, M., & López, C. C. (2025). The Health Impact of Cocoa from Cultivation to the Formation of Biogenic Amines: An Updated Review. *Foods*, 14(2), 255. <https://doi.org/10.3390/foods14020255>
- Polania Bello, V. (2023). Diagnosis of the agronomic and phytosanitary management practices of cocoa producers in San José del Fragua: the case of the Cerafín Garcia property. *Environmental Research and Ecotoxicity*, 2, 55. <https://doi.org/10.56294/ere202355>
- Puig, A. S., Irish, B., Ayala-Silva, T., Wurzel, S., & Gutierrez, O. (2022). Effect of Cacao Black Pod Rot Screening Method on Disease Reaction Determination. *The 1st International Online Conference on Agriculture&mdash;Advances in Agricultural Science and Technology*, 71. <https://doi.org/10.3390/IOCAG2022-12215>
- Quintero, I., Ceccaldi, A., Martínez, H., Santander, M., Rodríguez, J., & Escobar, S. (2025). Dry cacao pulp in chocolate bars: A sustainable, nutrient-rich sweetener with enhanced sensory quality through refractance windows drying. *Applied Food Research*, 5(1), 100700. <https://doi.org/10.1016/j.afres.2025.100700>
- Raj, M., & Prahadeeswaran, M. (2025). Revolutionizing agriculture: a review of smart farming technologies for a sustainable future. *Discover Applied Sciences*, 7(9), 937. <https://doi.org/10.1007/s42452-025-07561-6>
- Ray, R. K., Chakravarty, S., Dash, S., Ghosh, A., Mohanty, S. N., Reddy Chirra, V. R., Ayouni, S., & Khan, M. I.

- (2025). Precision pest management in agriculture using Inception V3 and EfficientNet B4: A deep learning approach for crop protection. *Information Processing in Agriculture*.  
<https://doi.org/10.1016/j.inpa.2025.09.005>
- Schmidt, J. E., Puig, A. S., DuVal, A. E., & Pfeufer, E. E. (2023). Phyllosphere microbial diversity and specific taxa mediate within-cultivar resistance to *Phytophthora palmivora* in cacao. *MSphere*, 8(5).  
<https://doi.org/10.1128/msphere.00013-23>
- Shafay, M., Hassan, T., Owais, M., Hussain, I., Khawaja, S. G., Seneviratne, L., & Werghi, N. (2025). Recent advances in plant disease detection: challenges and opportunities. *Plant Methods*, 21(1), 140.  
<https://doi.org/10.1186/s13007-025-01450-0>
- Song, X., Yan, L., Liu, S., Gao, T., Han, L., Jiang, X., Jin, H., & Zhu, Y. (2025). Agricultural Image Processing: Challenges, Advances, and Future Trends. *Applied Sciences*, 15(16), 9206.  
<https://doi.org/10.3390/app15169206>
- Taha, M. F., Mao, H., Zhang, Z., Elmasry, G., Awad, M. A., Abdalla, A., Mousa, S., Elwakeel, A. E., & Elsherbiny, O. (2025). Emerging Technologies for Precision Crop Management Towards Agriculture 5.0: A Comprehensive Overview. *Agriculture*, 15(6), 582. <https://doi.org/10.3390/agriculture15060582>
- Tan, M., & Le, Q. V. (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. Cornell University. <https://doi.org/https://doi.org/10.48550/arXiv.1905.11946>
- Villalobos-Culqui, C., Valles-Coral, M. A., García-Rivas-Plata, C., & Tuesta-Hidalgo, O. A. (2025). Modelo de visión artificial basada en redes neuronales convolucionales para identificación de mazorca negra en plantaciones de cacao. *Revista Científica de Sistemas e Informática*, 5(1), e678.  
<https://doi.org/10.51252/rcsi.v5i1.678>
- Vinci, G., Ruggeri, M., Gobbi, L., & Savastano, M. (2024). Social Life Cycle Assessment of Cocoa Production: Evidence from Ivory Coast and Ghana. *Resources*, 13(10), 141.  
<https://doi.org/10.3390/resources13100141>
- Wang, Y., Deng, Y., Zheng, Y., Chattopadhyay, P., & Wang, L. (2025). Vision Transformers for Image Classification: A Comparative Survey. *Technologies*, 13(1), 32.  
<https://doi.org/10.3390/technologies13010032>
- Waqas, M., Naseem, A., Humphries, U. W., Hlaing, P. T., Dechpichai, P., & Wangwongchai, A. (2025). Applications of machine learning and deep learning in agriculture: A comprehensive review. *Green Technologies and Sustainability*, 3(3), 100199. <https://doi.org/10.1016/j.grets.2025.100199>
- Wu, K., Ji, Z., Wang, H., Shao, X., Li, H., Zhang, W., Kong, W., Xia, J., & Bao, X. (2025). A Comprehensive Review of AI Methods in Agri-Food Engineering: Applications, Challenges, and Future Directions. *Electronics*, 14(20), 3994. <https://doi.org/10.3390/electronics14203994>
- Zahlul Ikhsan, Muhammad Al-Ikhlās, Yaherwandi, Hasmiandy Hamid, Aulia Oktavia, & Dandy Ahamefula Osibe. (2024). Level of Pest Infestation on Cocoa (*Theobroma cacao* L.) Variety BL-50 in Tanah Datar Regency, West Sumatera Province, Indonesia. *Andalasian International Journal of Entomology*, 2(1), 38–47. <https://doi.org/10.25077/aijent.2.1.38-47.2024>