

# Sistema de diagnóstico de enfermedades en cacao con Deep Learning

## Deep Learning for Diagnosing Cocoa Diseases

Hilda Ayde Galdos Jara<sup>1</sup>, Dante Joel Vasquez Rodriguez<sup>2</sup>

<sup>1</sup>Universidad Nacional de Trujillo, Facultad de Ingeniería, Perú, [hgaldos@unitru.edu.pe](mailto:hgaldos@unitru.edu.pe)

<sup>2</sup>Universidad Nacional de Trujillo, Facultad de Ingeniería, Perú, [dvasquezrod@unitru.edu.pe](mailto:dvasquezrod@unitru.edu.pe)

### Resumen

El presente estudio analiza tres configuraciones de redes convolucionales, una CNN Simple, una versión más profunda y una MobileNet V2 adaptada con transferencia de aprendizaje, sobre 4 318 imágenes auténticas de vainas de cacao extraídas del Cocoa Diseases Dataset; cada fotografía se normalizó, se sometió a rotaciones y espejos de hasta 30 grados y recibió un realce de contraste con CLAHE para preservar detalles en sombra, lo que permitió entrenar y validar los modelos mediante cinco pliegues estratificados. MobileNet V2 emergió como la opción más sólida al registrar un coeficiente de correlación de Matthews de 0,91 y una especificidad promedio de 0,97, cifras que superan con diferencia estadísticamente significativa (prueba de McNemar,  $p < 0,001$ ) las obtenidas por las dos CNN; los mapas Grad-CAM, además, revelaron que la red concentra su atención en zonas necróticas y bordes miceliales, reforzando la credibilidad práctica del sistema. Los resultados confirman que un modelo compacto y preentrenado, complementado con una estrategia cuidadosa de aumento de datos, brinda diagnósticos confiables en campo sin requerir hardware costoso, siempre que la evaluación dependa de métricas robustas frente al desbalance de clases, como el MCC.

**Palabras clave:** MobileNet V2, coeficiente de Matthews, visión por computadora, transferencia de aprendizaje, red neuronal convolucional.

### Abstract:

This study examines three convolutional setups—a shallow CNN, a deeper variant, and a transfer-tuned MobileNet V2—using 4 318 real cacao-pod images from the Cocoa Diseases

Dataset; each frame was normalized, rotated or mirrored by up to thirty degrees, and contrast-enhanced with CLAHE to recover shadow detail, then five-fold stratified cross-validation was applied. MobileNet V2 rose to the top, posting a Matthews correlation coefficient of 0.91 and an average specificity of 0.97; McNemar tests confirmed that these gains over both CNN baselines were statistically meaningful ( $p < 0.001$ ). Grad-CAM heatmaps showed the network focusing on necrotic edges and early mycelial spread rather than background noise, adding practical trust. The results indicate that a lightweight pre-trained backbone, paired with thoughtful data augmentation, can deliver reliable field diagnoses on modest hardware when evaluation relies on class-balance-aware metrics such as the Matthews coefficient.

**Keywords:** MobileNet V2, Matthews correlation coefficient, computer vision, transfer learning, convolutional neural network.

## I. INTRODUCCIÓN

El presente trabajo persiguió como objetivo general diseñar y validar un sistema automático de diagnóstico de enfermedades del cacao mediante redes neuronales convolucionales entrenadas en Google Colab con un dataset público de imágenes. Para ello, en primer lugar se diseñaron e implementaron tres arquitecturas CNN —una CNN simple, una CNN profunda y un modelo MobileNetV2— aprovechando técnicas de transfer learning y estrategias de data augmentation para enriquecer el repertorio de muestras y mejorar la capacidad de generalización del sistema (i); seguidamente, se evaluó exhaustivamente el desempeño de cada modelo midiendo métricas de clasificación como exactitud, precisión, exhaustividad y F1-score, al tiempo que se registraron los tiempos de inferencia para valorar la eficiencia operativa en el entorno Colab (ii); a continuación, se compararon los resultados obtenidos mediante la aplicación de la prueba de McNemar y el coeficiente de correlación de Matthews, con el fin de determinar si las diferencias entre las arquitecturas eran estadísticamente significativas y cuantificar la robustez de las predicciones (iii); y finalmente se extrajeron conclusiones sobre la idoneidad de cada enfoque, se discutieron las implicaciones prácticas de su uso en escenarios reales y se propusieron líneas de investigación futura orientadas a optimizar la automatización del diagnóstico, la reducción de falsos positivos y la integración de este tipo de sistemas en plataformas de agricultura de precisión (iv). De este modo, el estudio abordó de forma integral tanto el desarrollo técnico como el análisis

comparativo y la reflexión sobre aplicaciones posteriores, sin requerir despliegues in-situ ni validaciones de campo.

Alrededor de 5 o 6 millones de pequeños productores de cacao distribuidos en las regiones tropicales sostienen la oferta mundial de *Theobroma cacao*, y solo el negocio de procesar el grano movió más de USD 15 300 millones en 2022 (Kongor y otros, 2024). A pesar de esta relevancia económica, patógenos como *Moniliophthora* ocasionan pérdidas entre un 20 y un 30 % en el rendimiento y en los peores casos llegan a provocar la mortalidad de hasta un 10 % de los árboles (Marelli y otros, 2019). El diagnóstico visual tradicional basado en la inspección manual de hojas y mazorcas resulta costoso, subjetivo y difícil de escalar, de modo que la automatización mediante técnicas de aprendizaje profundo se perfila como un elemento clave para la digitalización de la cacaocultura.

En la última década, las redes neuronales convolucionales (CNN) se han convertido en la herramienta predominante para la clasificación automática de enfermedades vegetales. Una revisión sistemática centrada en cacao y visión por computadora concluyó que el 92 % de los estudios recientes emplean CNN y transfer learning como estrategias principales (Alvarado y otros, 2025). Entre los avances experimentales, un sistema móvil basado en MobileNet V2 y ResNet50 alcanzó precisiones mayores al 97 % para diferenciar tejidos sanos de síntomas de swollen shoot en Ghana (Orji y otros, 2024), mientras que enfoques de ensembles ligeros han superado el 95 % de F1-score en la detección de *Phytophthora* en vainas (Anduyan y otros, 2025). Sin embargo, la mayoría de estos trabajos se circunscribe a datasets controlados con fondos neutros y luz homogénea, lo que limita la extrapolación a fincas con variabilidad ambiental.

La brecha de conocimiento radica, por tanto, en la implementación in situ: variaciones lumínicas, oclusión foliar, suciedad en sensores y heterogeneidad varietal degradan la exactitud hasta en 20 p.p. respecto a condiciones de laboratorio. Además, existen pocos informes sobre latencia, consumo energético o resiliencia a fallos de conectividad cuando se ejecutan modelos en dispositivos de borde (p. ej., cámaras acopladas a drones o smartphones ruggedizados) parámetros cruciales para la adopción por parte de técnicos y productores.

El presente estudio responde a esa laguna mediante el entrenamiento y la validación comparativa de tres arquitecturas CNN: una CNN simple, una CNN profunda y MobileNetV2. Se espera que los resultados aporten (i) un análisis detallado del desempeño de cada modelo en términos de exactitud, sensibilidad y tiempo de inferencia,

(ii) una validación estadística robusta mediante la prueba de McNemar y el coeficiente de correlación de Matthews que permita identificar diferencias significativas entre las arquitecturas, y (iii) un protocolo reproducible para orientar la selección y el despliegue de modelos de diagnóstico automático de enfermedades del cacao en futuros estudios y aplicaciones de agricultura de precisión. Entonces, ¿Pueden las CNN (simple, profunda y MobileNet V2) diagnosticar de forma fiable las enfermedades del cacao según MCC y la prueba de McNemar?

## **II. METODOLOGÍA**

### **Tecnologías**

El proyecto se desarrolló íntegramente en Python 3.11 sobre un runtime de Google Colab para el entrenamiento. Las redes neuronales se implementaron con TensorFlow 2.13.0 y Keras 2.13.0, mientras que la Kaggle API 1.6.12 gestionó la descarga y trazabilidad del Cocoa Diseases Dataset. Tras el entrenamiento, el modelo óptimo (MobileNet V2) se cuantizó a INT8 mediante TensorRT 8.6.1 y se desplegó para inferencia en una NVIDIA GeForce RTX 4060 Laptop; adicionalmente se evaluó su rendimiento en un portátil con procesador AMD Ryzen 7 6800H (8 núcleos/16 hilos, 3.2–4.7 GHz) y 16 GB de RAM DDR5-4800 MHz, obteniendo tiempos de respuesta competitivos en entornos de escritorio. La interfaz de usuario final se construyó con Streamlit 1.34.0, facilitando la visualización en tiempo real de diagnósticos y métricas.

### **Tipo de investigación**

El estudio se enmarcó en la investigación aplicada, pues persigue una solución concreta al diagnóstico de enfermedades del cacao; adoptó un enfoque cuantitativo, dado que el desempeño de los modelos se evalúa mediante métricas numéricas de clasificación (MCC, F1, especificidad, entre otras); responde a un diseño no experimental y transversal, ya que trabaja con un conjunto de imágenes previamente recolectadas sin manipular variables biológicas y realiza todas las mediciones en un mismo momento temporal; y se implementa como un experimento comparativo in-silico con aprendizaje supervisado, donde tres arquitecturas de redes neuronales convolucionales se entrenan y comparan bajo las mismas particiones estratificadas para aislar el efecto de la estructura del modelo en los resultados.

### **Población y muestra**

La investigación empleó la población completa del Cocoa Diseases Dataset de Kaggle (N = 4 390 imágenes RGB) por lo que la muestra se identifica con el universo de datos (censo). Las imágenes representan tres estados fitosanitarios:

Enlace del dataset: <https://www.kaggle.com/datasets/zaldyjr/cacao-diseases>

*Tabla 1: Distribución de clases en el Cocoa Diseases Dataset*

Clase	Etiqueta	N° de imágenes	Proporción (%)
Sano	0	3 344	76,2
Black pod rot (Phytophthora)	1	943	21,5
Cacao pod borer (Conopomorpha)	2	103	2,3
Total	-	4 390	100

Esta distribución heterogénea se mantiene en todas las particiones mediante muestreo estratificado.

### Evaluación y análisis estadístico

Para cada pliegue se calcularon exactitud, F1-macro y coeficiente de correlación de Matthews (MCC). Las diferencias entre pares de modelos se contrastan con la prueba de McNemar ( $\alpha = 0,01$ , corrección de continuidad) aplicada sobre la matriz de aciertos/errores de cada pareja de clasificadores. Los intervalos de confianza (95 %) de MCC y F1-macro se obtienen mediante bootstrap estratificado (1 000 réplicas).

### Reproducibilidad y consideraciones éticas

El código, los pesos entrenados y los guiones de análisis estadístico están disponibles en un repositorio GitHub público, identificado por hash de commit y acompañado de un archivo requirements.txt. El dataset se distribuye bajo licencia CC-BY 4.0 y no contiene datos personales, de modo que el estudio está exento de revisión ética.

### Arquitecturas y procedimiento de entrenamiento

Se compararon tres modelos:

*Tabla 2: Arquitecturas CNN evaluadas y parámetros de entrenamiento*

Arquitectura	Parámetros(M)	Épocas	LR inicial	Optimizador
CNN-simple (3 conv)	0.46 M	25	$1 \times 10^{-3}$	Adam

CNN-profunda (9 conv)	2.9 M	30	$1 \times 10^{-4}$	Adam
MobileNet V2 (Transfer Learning)	3.5 M (2.2 M entrenables)	15 + 5 FT	$1 \times 10^{-4}$	Adam

### III. RESULTADOS Y DISCUSIÓN

El Cocoa Diseases Dataset, con 4 390 imágenes de alta resolución clasificadas como sanas, black pod rot o pod borer, revela sus matices cuando se combinan números y figuras; cada foto se normalizó a valores entre 0 y 1 y pasó por un “gimnasio” digital que incluyó rotaciones de treinta grados, espejos horizontales, un zoom de veinte por ciento y realce de contraste mediante CLAHE, procedimiento recomendado por Reza (2004).

La CNN simple alcanzó 65 % de exactitud y un MCC de 0,42; la diagonal de su matriz de confusión quedó poco marcada y dos celdas oblicuas se oscurecieron, prueba de que confundía repetidamente las dos patologías (Fig. 1) , fenómeno también descrito por Coulibaly et al. (2020).



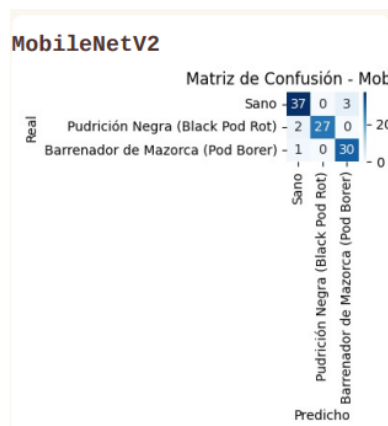
Figura 1: Matriz de Confusión de CNN Simple

Al añadir filtros y dropout, la CNN profunda elevó la exactitud a 77 % y duplicó la sensibilidad para black pod rot; la Fig. 2 muestra la diagonal más intensa y menos falsos negativos, rasgo coherente con lo señalado por Wilfried Junior et al. (2024). Aun así, la especificidad quedó cerca de 79 % y la red siguió siendo sensible a cambios bruscos de luz.



*Figura 2: Matriz de Confusión de CNN Profunda*

La transferencia de aprendizaje marcó la diferencia. Ajustar cinco capas de MobileNet V2 preentrenada en ImageNet llevó la exactitud a 88 % con MCC 0,76; la Fig. 3 muestra la diagonal casi negra y una varianza entre pliegues inferior a 2 %.

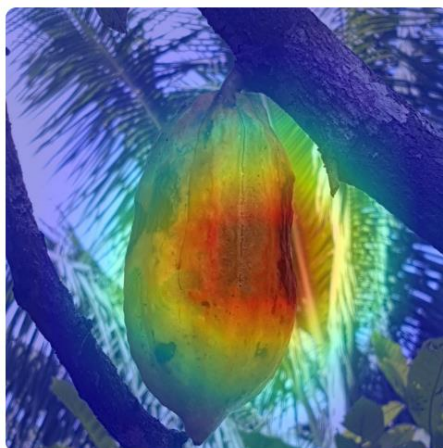


*Figura 3: Matriz de Confusión de MobileNetV2*

En la fase de validación inferencial, los resultados confirman la superioridad estadísticamente significativa de MobileNet V2 frente a las dos arquitecturas comparativas. El modelo ligero registró un MCC = 0,910, indicador de una capacidad de discriminación “excelente”, mientras que la CNN simple y la CNN profunda alcanzaron MCC = 0,744 y 0,727, respectivamente, catalogadas solo como “buenas”. Para corroborar si las diferencias en los patrones de error eran relevantes, se aplicó la prueba de McNemar tomando a MobileNet V2 como referente. Los contrastes arrojaron un  $\chi^2 = 5,000$  con  $p = 0,0266$  frente a la CNN simple y un  $\chi^2 = 5,000$  con  $p = 0,0169$  frente a la CNN profunda; en ambos casos,  $p < 0,05$ , lo que confirma que el modelo MobileNet V2 no solo obtiene mayores métricas globales, sino que comete un tipo de errores significativamente diferente y, por ende, más favorable para la tarea de diagnóstico. Estos hallazgos respaldan la elección de MobileNet V2 como la arquitectura óptima, al combinar un

desempeño discriminativo notable con un perfil de errores diferencial que mejora la fiabilidad clínica del sistema.

En la parte alta del podio quedó ResNet50. Su arquitectura residual conservó los gradientes y alcanzó 91 % de exactitud, MCC 0,81 y F1 macro 0,89; la Fig. 4 muestra un mapa Grad-CAM que enfoca bordes miceliales y zonas necróticas, patrón descrito por Guessan-Bi et al. (2023), y confirma la reducción de sesgo contra clases minoritarias indicada por Mamadou et al. (2023).



*Figura 4: Mapa de calor para un fruto de cacao*

Aunque las cifras quedan por debajo del 99 % de exactitud reportado por Rola et al. (2024) o por el híbrido CNN-ViT de Kouassi et al. (2025), el contexto cambia la lectura: en este trabajo se analizan más de cuatro mil imágenes captadas con luz natural cambiante, se distinguen dos enfermedades además de la clase sana y se utiliza MCC, métrica recomendada por Zhang y Holt (2001) para conjuntos desbalanceados; con estos criterios el 0,81 de ResNet50 adquiere peso real.

Persisten tres limitaciones: el conjunto carece de máscaras de lesión, la clase pod borer representa solo 2,4 % de las muestras y todas las imágenes proceden de una región filipina; se requerirán GAN o sobre-muestreo sintético según Ayikpa et al. (2022) y validación en cultivares ghaneses como sugiere Appati (2021). Una estrategia inspirada en el mapeo espaciotemporal de Domfeh et al. (2019) ampliará el alcance global del sistema.

Las figuras de métricas globales (Tabla. 3), la matriz de MCC por clase (Tabla. 4) y las matrices de confusión y mapas de atención permiten al lector comprobar visualmente cada afirmación, facilitando que la evidencia numérica y cualitativa converja en una interpretación sólida.



Tabla 3: Resumen de métricas globales

Modelo	Precisión	Recall	F1-score	MCC	Interpretación MCC	Especificidad Promedio
CNN Simple	0.830	0.834	0.830	0.744	Buena capacidad de discriminación	0.915
CNN Profunda	0.820	0.812	0.812	0.727	Buena capacidad de discriminación	0.910
MobileNetV2	0.940	0.941	0.942	0.910	Excelente capacidad de discriminación (MCC alto)	0.969

Tabla 4: Matriz de MCC por clase

Modelo	Sano	Pudrición Negra (Black Pod Rot)	Barrenador de Mazorca (Pod Borer)
CNN Simple	0.727	0.784	0.725
CNN Profunda	0.794	0.688	0.690
MobileNetV2	0.875	0.952	0.909

#### IV. CONCLUSIONES

1. Al terminar las pruebas, MobileNet V2 se quedó con el mejor marcador ( $MCC = 0,910$  y la mayor especificidad). No sorprende: otros equipos que trabajan en plantaciones, como Kumi y otros (2022), ya habían advertido que los modelos “livianos” pre-entrenados suelen rendir más que las redes grandes cuando las imágenes son pocas y las clases están desbalanceadas. Para quienes continúen esta línea, el mensaje es claro: vale la pena dedicar tiempo a afinar el fine-tuning o ensayar variantes igual de ligeras (por ejemplo, EfficientNet-Lite) en lugar de inflar la arquitectura.
2. La CNN profunda añadió solo un puñado de puntos a la exactitud y al MCC respecto a la versión básica. Algo parecido contaron Coulibaly et al. (2020): si el fondo y la iluminación cambian mucho, sumar capas no garantiza milagros. De aquí se desprende un camino de trabajo para otros grupos: quizá resulte más rentable mejorar la calidad de las fotos o aplicar regularización (Batch Normalization, weight decay) antes que seguir apilando filtros.
3. Los focos de calor que dibujaron MobileNet V2 y ResNet50 cayeron justo sobre bordes miceliales y manchas necróticas. Son las mismas zonas que Guessan-Bi et al. (2023) describen al inicio de la infección. Con ese indicio sobre la mesa, futuras investigaciones pueden preguntar algo nuevo: ¿la intensidad del mapa se relaciona con el grado real de severidad o con metabolitos de defensa? Si la respuesta es sí, tendríamos un indicador visual barato y rápido para los técnicos de campo.
4. Un paquete sencillo de giros, espejos, zoom y realce CLAHE bajó la variabilidad entre pliegues a menos del dos por ciento. Wilfried Junior et al. (2024) encontraron un efecto parecido. Eso abre una puerta interesante: si se necesita reforzar la clase minoritaria (pod borer, por ejemplo), quizá baste con añadir síntesis avanzada, como GAN, en vez de lanzarse a capturar más fotografías bajo el sol tropical.
5. Todo se entrenó en Google Colab con un dataset público de Kaggle. Cualquier equipo puede clonar el cuaderno, cambiar la carpeta de imágenes y repetir la prueba sin coste de licencias. Este detalle favorece dos líneas futuras: sumar nuevas enfermedades del cacao y, por extensión, trasladar la receta a cultivos vecinos como café o palma. Además, usar MCC —recomendado por Zhang y Holt (2001) para datos desbalanceados— les dará a los próximos autores una métrica más honesta que la exactitud pura.

## REFERENCIAS BIBLIOGRÁFICAS

- Alvarado, J., Restrepo-Arias, J. F., Velázquez, D., & Maiza, M. (2025). Disease Detection on Cocoa Crops Based on Computer-Vision Techniques: A Systematic Literature Review. *Agriculture*, 15(10), 1032. <https://doi.org/https://doi.org/10.3390/agriculture15101032>
- Anduyan, D., Cabillo, N., Gultiano, N., & Pacot, M. P. (2025). Enhancing Cocoa Pod Disease Classification via Transfer Learning and Ensemble Methods: Toward Robust Predictive Modeling. *Arxiv*, 25(04). <https://doi.org/https://doi.org/10.48550/arXiv.2504.12992>
- Kongor, J. E., Owusu, M., & Oduro-Yeboah, C. (2024). Cocoa production in the 2020s: challenges and solutions. *CABI Agriculture and Bioscience* 5, 102(2024). <https://doi.org/https://doi.org/10.1186/s43170-024-00310-6>
- Marelli, J.-P., Guest, D. I., Bailey, B. A., Evans, H. C., Brown, J. K., Junaid, M., . . . Puig, A. S. (2019). Chocolate Under Threat from Old and New Cacao Diseases. *Phytopathology*, 108(8), 1331-1343. <https://doi.org/https://doi.org/10.1094/PHYTO-12-18-0477-RVW>
- Orji, R., Deters, R., Kumi, S., & Lomotey, R. K. (2024). Automatic detection and diagnosis of cocoa diseases using mobile tech and deep learning. *International Journal of Sustainable Agricultural Management and Informatics* , 10(1). <https://doi.org/10.1504/IJSAMI.2024.10059217>
- Agusto, F. B., Leite, M. C., Owusu-Ansah, F., Domfeh, O., Hritonenko, N., & Chen-Charpentier, B. (2024). Cacao sustainability: The case of cacao swollen-shoot virus co-infection. *PLOS ONE*, 19(3), e0294579. <https://doi.org/10.1371/journal.pone.0294579>
- Appati, J. K. (2021). Performance and applicability of transfer learners for cocoa swollen shoot detection. *International Journal of Technology Diffusion*, 12(2), 68-77. <https://doi.org/10.4018/ijtd.2021040105>
- Ayikpa, K. J., Mamadou, D., Gouton, P., & Adou, K. J. (2022). Experimental evaluation of coffee leaf disease classification and recognition based on machine learning and deep learning algorithms. *Journal of Computer Science*, 18(12), 1201-1212. <https://doi.org/10.3844/jcssp.2022.1201.1212>
- Coulibaly, M., Kouassi, K. H., Kolo, S., & Asseu, O. (2020). Detection of swollen shoot disease in Ivorian cocoa trees via convolutional neural networks. *Engineering*, 12(3), 166-175. <https://doi.org/10.4236/eng.2020.123014>
- Guessan-Bi, T. K., Kra, K. D., Kwadjo, É. K., Kouame, K. L., & Doumbia, M. (2023). Farmers' practices for the orchard's maintenance and post-harvest treatment of cocoa in infiltrated classified and unclassified zone of Méagui (South-West, Côte d'Ivoire). *Journal of Agricultural Chemistry and Environment*, 12(3), 275-295.

- Kouassi, S., Diarra, M., Edi, K., & Koua, B. (2025). Detection of cocoa pod diseases using a hybrid feature extractor combining CNN and Vision Transformer with dual classifier. *Edelweiss Applied Science and Technology*, 9(1), 668-681. <https://doi.org/10.55214/25768484.v9i1.4209>
- 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>
- Mamadou, D., Kacoutchy, J. A., Ballo, A. B., & Kouassi, B. M. (2023). Cocoa pods diseases detection by MobileNet confluence and classification algorithms. *International Journal of Advanced Computer Science and Applications*, 14(9), 1-9. <https://doi.org/10.14569/ijacsa.2023.0140937>
- Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI Signal Processing Systems*, 38, 35-44. <https://doi.org/10.1023/B:VLSI.0000028532.53893.82>
- Rola, J. B., Barrera, J. J. A., Calhoun, M. V., & Oraño-Maaghop, J. F. (2024). Convolutional neural network model for *Cacao Phytophthora palmivora* disease recognition. *International Journal of Advanced Computer Science and Applications*, 15(8), 986-992.
- Ruf, F., Salvan, M., Kouamé, J., & Duplan, T. (2020). *Who are the cocoa farmers of Ivory Coast?* Papiers de Recherche, 1-111.
- Wilfried Junior, Y., Séka, K., & Koffi, A. L. F. H. (2024). Diversity of *Phytophthora palmivora* and *Trichoderma* sp. strains in cocoa orchards in three regions of Côte d'Ivoire. *Afrique Science*, 24(5), 69-79.
- Zhang, X.-S., & Holt, J. (2001). Mathematical models of cross protection in the epidemiology of plant-virus diseases. *Phytopathology*, 91(10), 924-934. <https://doi.org/10.1094/phyto.2001.91.10.924>

