Rev. Cient. Sist. Inform. 6(1), e1385, doi: 10.51252/rcsi.v6i1.1385

UNIVERSIDAD NACIONAL DE SAN MARTÍN

RCSI
https://revistas.unsm.edu.pe/index.php/rcsi
e-ISSN: 2709-992X

Original article
Artículo original
Jan-Jun, 2026

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

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

Navarro-Cabrera, Jorge Raul[1]*
Beraún-Barrantes, José Guillermo[1]
Cárdenas-García, Ángel[2]
Lozano-Carranza, Carlos Mauricio[3]

[1]Universidad de Huánuco, Huánuco, Peru
[2]Universidad Nacional de San Martín, Tarapoto, Peru
[3]TUSAN Ingenieros Consultores, Tarapoto, Peru

Received: 16 Sep. 2025 | **Accepted:** 30 Dec. 2025 | **Published:** 20 Jan. 2026

Corresponding author*: jorge.navarro@udh.edu.pe

How to cite this article: Navarro-Cabrera, J. R., Beraún-Barrantes, J. G., Cárdenas-García, A. & Lozano-Carranza, C. M. (2026). Application of deep learning in the phytosanitary detection of cocoa using computer vision. *Revista Científica de Sistemas e Informática, 6*(1), e1385. https://doi.org/10.51252/rcsi.v6i1.1385

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

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

© The authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Cocoa (Theobroma cacao L.) is one of the most important agricultural crops worldwide, playing a key role in the economies of tropical countries and in generating income for millions of smallholder farmers (Charry et al., 2025; Vinci et al., 2024). Its fermented and dried beans constitute the essential raw material for the chocolate industry and various confectionery products (Paparella et al., 2025; Quintero et al., 2025). However, cocoa production faces serious threats from fungal diseases and pests that drastically reduce yields and compromise bean quality, affecting the global competitiveness of the value chain (Cilas & Bastide, 2020; Delgado-Ospina et al., 2021; Zahlul Ikhsan et al., 2024). Among the most relevant pathologies are Black Pod Rot, caused by Phytophthora spp., and the Pod Borer, both responsible for significant economic losses in tropical production areas (Magfirah et al., 2025; Puig et al., 2022; Schmidt et al., 2023).

Traditionally, the detection of these conditions has relied on visual inspections carried out by farmers or specialists, a subjective procedure that depends on individual experience and the morphological interpretation of symptoms (Miyittah et al., 2022; Polania Bello, 2023). This approach, in addition to being slow and costly, presents high variability in diagnostic accuracy, which delays the implementation of effective control measures and increases disease spread. Consequently, the agricultural industry demands more reliable, faster, and reproducible methods for the early identification of pests and diseases, enabling optimized management processes and improved crop productivity (Raj & Prahadeeswaran, 2025; Wu et al., 2025).

In this context, advances in computer vision and machine learning (ML) have opened new opportunities for automating diagnostic tasks in precision agriculture (Taha et al., 2025; Waqas et al., 2025). These technologies enable the analysis of large volumes of visual data through algorithms capable of recognizing complex patterns in images of leaves, fruits, or stems (Injante et al., 2025; Lebrini & Ayerdi Gotor, 2024). Artificial vision systems (AVS), by combining optical hardware and digital image processing software, have become non-destructive, fast, and cost-effective tools for the characterization and classification of agricultural products, achieving accuracy levels comparable to or even exceeding those obtained by human experts (Anjali et al., 2024; Song et al., 2025).

Deep learning (DL), as an evolution of ML, has revolutionized the field of computer vision by introducing models capable of learning hierarchical representations directly from data (Ray et al., 2025; Villalobos-Culqui et al., 2025). These approaches allow the automatic extraction of relevant image features, eliminating the need for manual feature engineering and increasing generalization capability in complex tasks (Mall et al., 2023). Owing to their multi-level architecture, DL models have demonstrated outstanding performance in the detection, segmentation, and classification of visual patterns, consolidating themselves as the dominant paradigm in applications such as agricultural diagnostics, quality control, and automated crop monitoring (Deepa et al., 2025; Shafay et al., 2025).

Nevertheless, the current literature reveals important limitations. Most studies focus on a single architecture or evaluate models without applying statistical methods that allow performance comparison with inferential significance. This lack of rigorous comparative analysis prevents a clear identification of which architectures are more stable, accurate, and efficient in agricultural contexts with limited computational resources. Furthermore, many experiments lack cross-validation schemes or reproducible protocols, hindering result replicability and practical adoption.

To address this gap, there is a need for systematic evaluations that compare contemporary deep learning approaches under a controlled experimental framework. These models involve different trade-offs between accuracy, complexity, and computational efficiency, making it essential to analyze their performance in automated classification tasks of agricultural pests and diseases. Comparative evaluation enables the establishment of technical guidelines for their implementation in intelligent monitoring systems and automated phytosanitary diagnostics, promoting solutions that combine high performance, stability, and operational feasibility in real production environments (Bono et al., 2026).

In this study, a computer vision approach driven by deep learning (Deep Learning–Driven Computer Vision) is proposed for the early and automatic detection of pests and diseases in cocoa fruits. A reproducible methodological process was implemented, integrating preprocessing strategies, stratified cross-validation, and statistical analysis to compare the performance of different deep learning models. The objective is to identify visual patterns associated with distinct phytosanitary conditions and to evaluate model stability and accuracy in automated classification scenarios. This work aims to provide technical and practical evidence for the development of intelligent agricultural diagnostic systems, promoting sustainability and efficiency in global cocoa production.

2. MATERIALS AND METHODS

The experimental development was structured as a reproducible computer vision pipeline designed for the automatic classification of diseases in cocoa fruits. The process comprised the following stages: (1) exploratory data analysis (EDA), (2) preprocessing, (3) training with stratified cross-validation, (4) evaluation using classification metrics, and (5) inferential statistical analysis to contrast model performance. Figure 1 presents the methodological framework, which ensures result traceability and comparability across configurations, enabling the replication of the study in different production contexts or with new datasets.

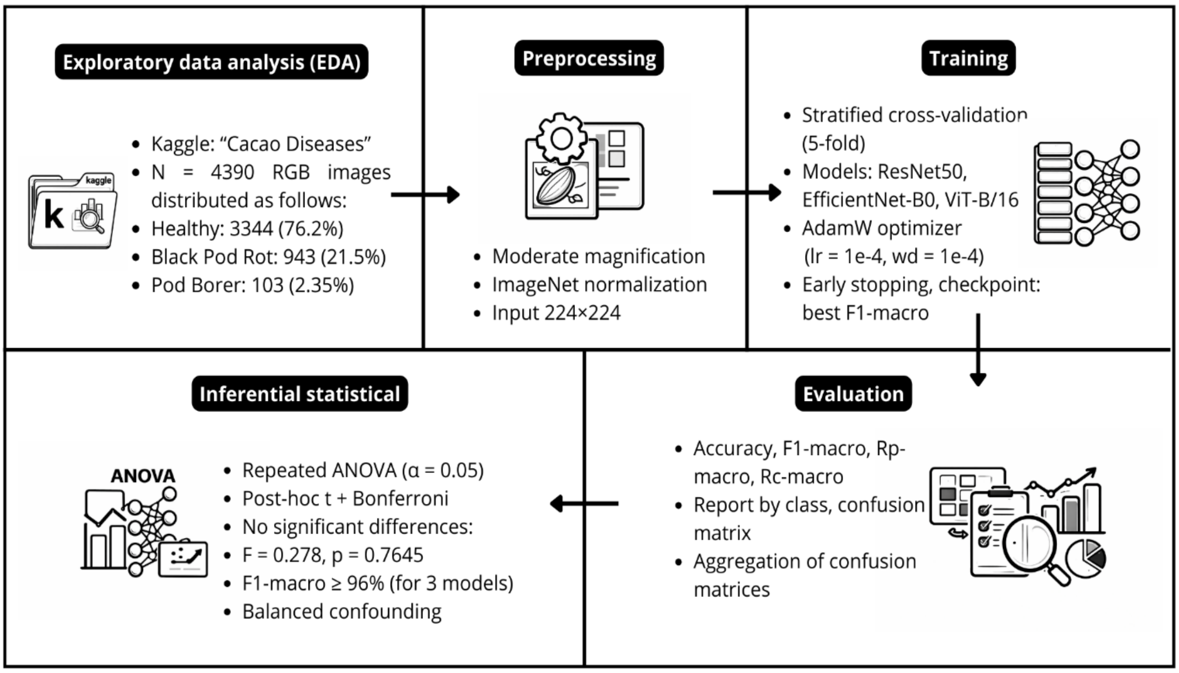

Figure 1. Proposed methodological framework

Dataset and Initial Preparation

The image dataset named "Cacao Diseases" was used, obtained from Kaggle (class-based folder structure). It consists of 4,390 RGB images of cacao fruits labeled into three categories: Black Pod Rot (943), Pod Borer (103), and Healthy (3,344). The images were organized using the ImageFolder structure (one folder per class) for direct consumption by PyTorch. Prior to training, a basic exploratory data analysis (EDA) was conducted: first, class counts were computed to identify imbalance; second, image size sampling was performed to estimate resolution variability; and finally, a visual mosaic of sample images per class was generated. Any corrupted or unreadable images were discarded.

Experimental Design and Data Partitioning

To obtain robust estimates and enable statistical comparison among architectures, five-fold stratified cross-validation was applied, ensuring the same class distribution in each fold. In every fold, each of the three models was trained and validated on identical data partitions, which subsequently allowed the fold to be treated as a "subject" in the repeated-measures analysis. A global random seed (42) was set to ensure reproducibility.

Preprocessing and Data Augmentation

In order to preserve chromatic signals relevant to symptoms such as lesions, mycelium, or discoloration, moderate data augmentation parameters were applied along with standard ImageNet normalization

Training:

```
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),
])
```

Validation:

```
eval_tf = transforms.Compose([
      transforms.Resize(int(img_size*1.15)),
      transforms.CenterCrop(img_size),
      transforms.ToTensor(),
      transforms.Normalize(IMAGENET_MEAN, IMAGENET_STD),
    ])
```

The input size was fixed at 224×224 pixels to ensure compatibility with the ImageNet standard and improve computational efficiency. Data augmentation transformations were kept conservative to avoid distorting fine lesion features.

Evaluated Models

Three deep learning architectures widely used in recent studies on agro-phytopathological image classification were evaluated: ResNet50 (He et al., 2015), EfficientNet-B0 (Tan & Le, 2020) and Vision Transformer (ViT-B/16) (Wang et al., 2025). All networks were pretrained on the ImageNet-1K dataset to leverage general visual representations learned from millions of natural images. Subsequently, the output layers were reconfigured to adapt the models to the specific problem domain. For ResNet50, the final layer was replaced with a fully connected layer containing three neurons corresponding to the target classes. In EfficientNet-B0, the last classifier layer was substituted with a three-class output, while in ViT-B/16, the original classification head was replaced by a linear layer with the same output dimensionality.

For all three models, full fine-tuning of parameters was performed rather than freezing the pretrained convolutional or attention layers. This methodological choice was motivated by the substantial difference between the source domain represented by ImageNet—based on natural objects and general scenes—and the target domain of this study, which focuses on images of cacao fruits affected by diseases and pests. Full weight adaptation allows the architectures to learn more discriminative representations specific to the phytopathological context, optimizing model sensitivity to subtle visual patterns and textures associated with different fruit health conditions.

Training Configuration

Model training was implemented using the PyTorch framework, applying a uniform set of hyperparameters to ensure comparability across architectures. The AdamW optimizer (Kingma & Ba, 2014) was used with an initial learning rate of 1×10^{-4} and a weight decay parameter of 1×10^{-4}, settings that balance gradient stability and model regularization during learning. The selected loss function was Cross-Entropy Loss, which is suitable for multiclass classification problems. The batch size was set to 32 images per iteration, and training was conducted for a maximum of 10 epochs per fold within the stratified cross-validation scheme.

To dynamically adjust the learning rate, a ReduceLROnPlateau scheduler was incorporated, reducing the learning rate by 50% when the validation loss failed to show significant improvement. In addition, an early stopping mechanism with a patience of three epochs was implemented, selecting as the optimal checkpoint the model corresponding to the highest macro F1 score achieved on the validation set. This combination of strategies helped control overfitting, stabilize convergence, and optimize computational resource usage, making the approach suitable for both hardware-constrained environments and edge inference scenarios.

Per-Fold Evaluation Protocol

At the end of each epoch, the model was evaluated on the corresponding fold's validation set, and once the early stopping criterion was triggered, the best model state was loaded to compute the final metrics for that fold: accuracy, macro F1 score, macro precision, and macro recall, as well as a per-class classification report (precision/recall/F1) and the confusion matrix.

Confusion matrices for each fold were stored and later aggregated per model in two ways: by summing absolute counts and by applying a row-normalized version (per-class recall). This approach enabled the analysis of systematic confusion patterns among classes without allowing a single fold to dominate the interpretation.

Statistical Comparison Between Models

To compare the mean performance of the three architectures, a repeated-measures statistical design was applied, considering the five folds generated during cross-validation. First, a long-format table was constructed including the fold (treated as the subject), the model (considered a within-subject factor with three levels), and the response metric, macro F1. Subsequently, a repeated-measures analysis of variance (RM-ANOVA) was performed, using the model as the within-subject effect and the fold as the observational unit, with a significance level set at $\alpha = 0.05$.

When statistical significance was detected, a post hoc analysis was conducted using paired t-tests between each pair of models, applying the Bonferroni correction to control the cumulative Type I error. This approach enabled a robust assessment of architecture-related differences while controlling for variability introduced by validation splits and avoiding conclusions dependent on a single train–validation partition.

3. RESULTS AND DISCUSSION

3.1. Exploratory data analysis (EDA)

After the initial exploration of the dataset, a total of 4,390 images were identified, distributed across three classes with a considerable imbalance. The Healthy category accounted for 3,344 images (76.2%), while Black Pod Rot included 943 images (21.5%), and Pod Borer only 103 images (2.35%), representing an approximate ratio of 32.5 between the majority and minority classes. A geometric analysis conducted on a sample of 300 images indicated that all captures exhibit an aspect ratio close to 1.0, with resolutions ranging from 1080×1080 to 2160×2160 pixels, confirming the uniformity of the acquisition format. These observations justified the use of stratified cross-validation to preserve class proportions within each fold, as well as the prioritization of the macro F1-score for checkpoint selection in order to mitigate bias induced by class imbalance. Additionally, moderate data augmentation techniques were applied during training to enhance the model's generalization capability without compromising the integrity of visual cues associated with lesion patterns (Figure 2).

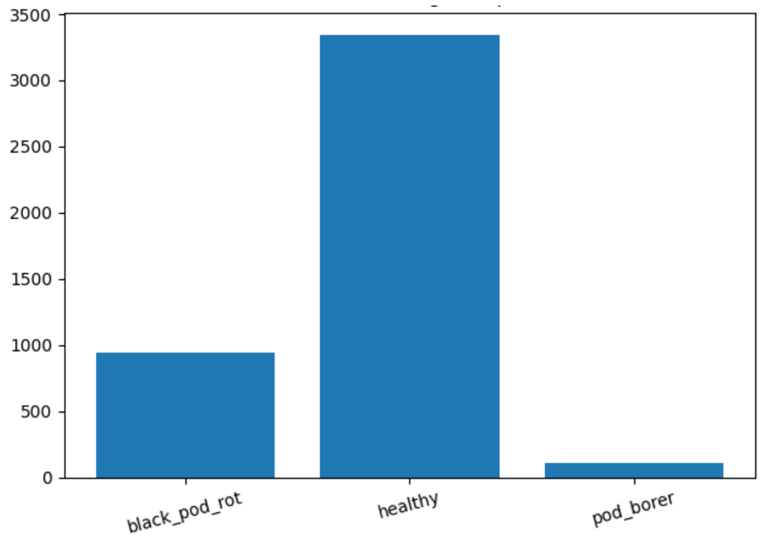

Figure 2. Distribution chart by classes

3.2. Overall performance of the models

The analysis of the models' overall performance revealed highly competitive results among the three evaluated architectures. The Vision Transformer (ViT-B/16) achieved the highest mean macro F1-score, with a value of 0.9697 and a standard deviation of 0.0114, along with an average accuracy of 0.9788. ResNet50 ranked second, with a mean macro F1-score of 0.9683, followed by EfficientNet-B0, which obtained 0.9666. The differences among the three models were minimal, with an average separation of approximately 0.3 percentage points between the best- and lowest-performing models in terms of the macro F1-score. This tightly clustered behavior reflects the robustness of the evaluated architectures and confirms the effectiveness of the preprocessing and validation scheme adopted in the study (Table 1).

Table 1. Summary of evaluation metrics by model

Model	Mean accuracy	Accuracy SD	Mean F1	SD F1	Mean Precision	Precision SD	Mean Recall	Recall SD
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. Training, stratified cross-validation, and inter-fold consistency

The evolution of the F1 macro during the training and validation process showed a stable convergence pattern in all three models, with no signs of overfitting and minimal fluctuations between folds. Figures 3, 4, and 5 show the consistency of the validation curves corresponding to EfficientNet-B0, ResNet50, and ViT-B/16, respectively, confirming the stability of the optimization process and the reproducibility of the results obtained.

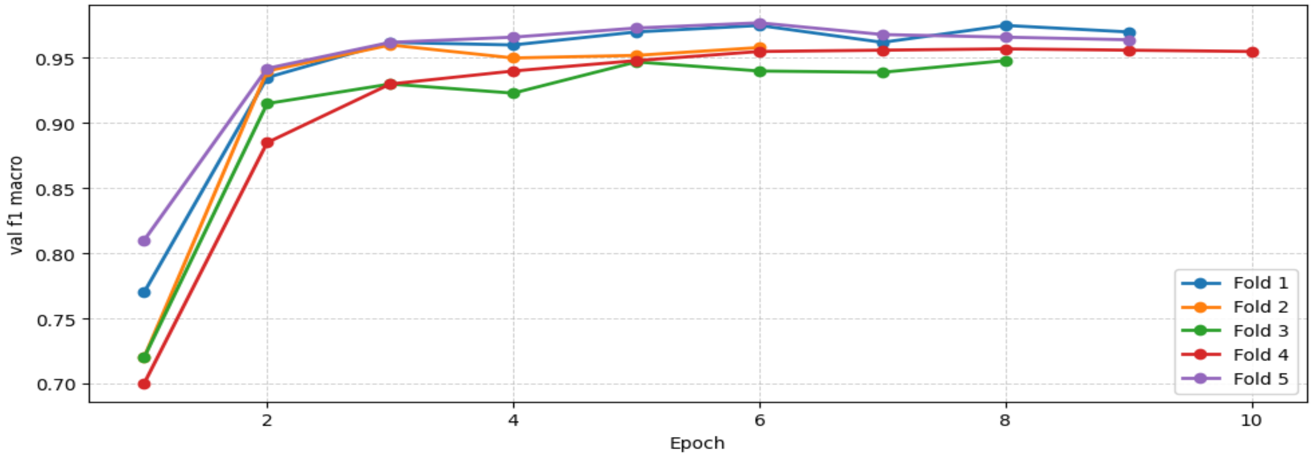

Figure 3. Evolution of the F1 Macro (validation) by fold in the EfficientNetB0 model

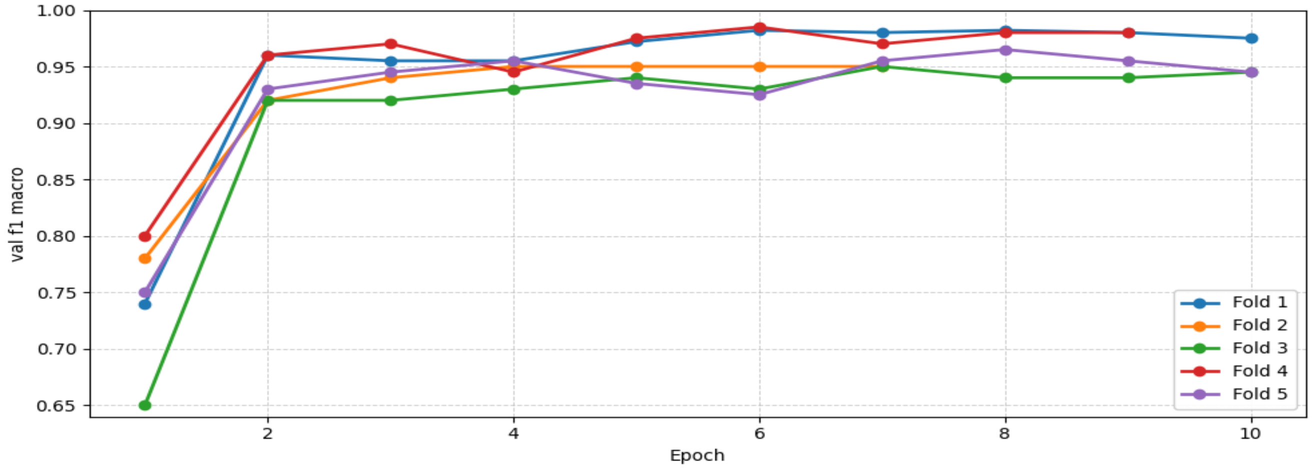

Figure 4. Evolution of the F1 Macro (validation) by fold in the ResNet50 model

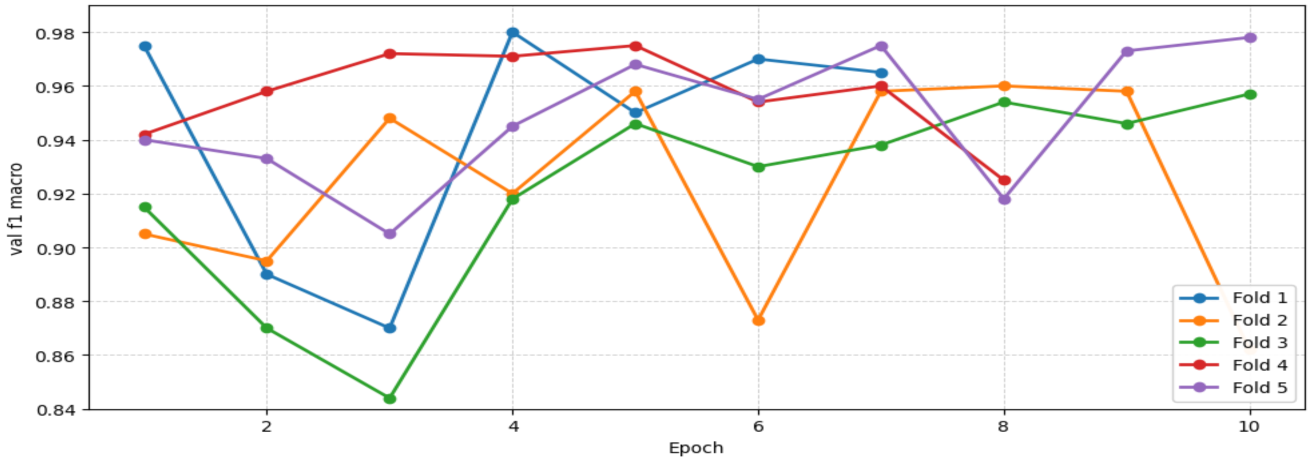

Figure 5. Evolution of the F1 Macro (validation) by fold in the ViT-B/16 model

The analysis of inter-fold consistency showed low variability across all experimental configurations, with standard deviations of the macro F1-score close to one hundredth, reflecting high stability in the training and validation process. This behavior suggests that the decisions adopted during image preprocessing and the use of stratified cross-validation contributed to the robustness of the procedure, preventing the influence of a dominant fold that could bias the overall averages. The detailed results for each architecture across the five folds are presented in Table 2, where the consistency of accuracy, precision, recall, and F1 metrics can be observed, evidencing balanced performance across models and data partitions.

Furthermore, the training and validation curves exhibited rapid convergence, reaching stability in fewer than ten epochs. The use of a learning rate scheduler facilitated loss stabilization during the final stages of training, while the early stopping criterion helped prevent overfitting. Taken together, these results confirm the reproducibility of the proposed pipeline and the inter-fold consistency of the models, supporting the validity of the conclusions derived from the comparative analysis.

Table 2. Stratified cross-validation metrics

Fold	Model	Accuracy	F1	Precision	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. Confusion matrices and behavior by class

The aggregated confusion matrices for each model, obtained from the sum of the five folds, along with their row-normalized versions, showed balanced behavior across classes, with no indication of dominant systematic confounding patterns. Row normalization allowed for a detailed analysis of recall by class, revealing slight variations between architectures that are consistent with the previously reported narrow range of macro F1 values. These variations do not alter the overall conclusion of the study: all three models consistently discriminate correctly between the Black Pod Rot, Pod Borer, and Healthy categories, maintaining homogeneous performance across the evaluated dataset (Figure 6).

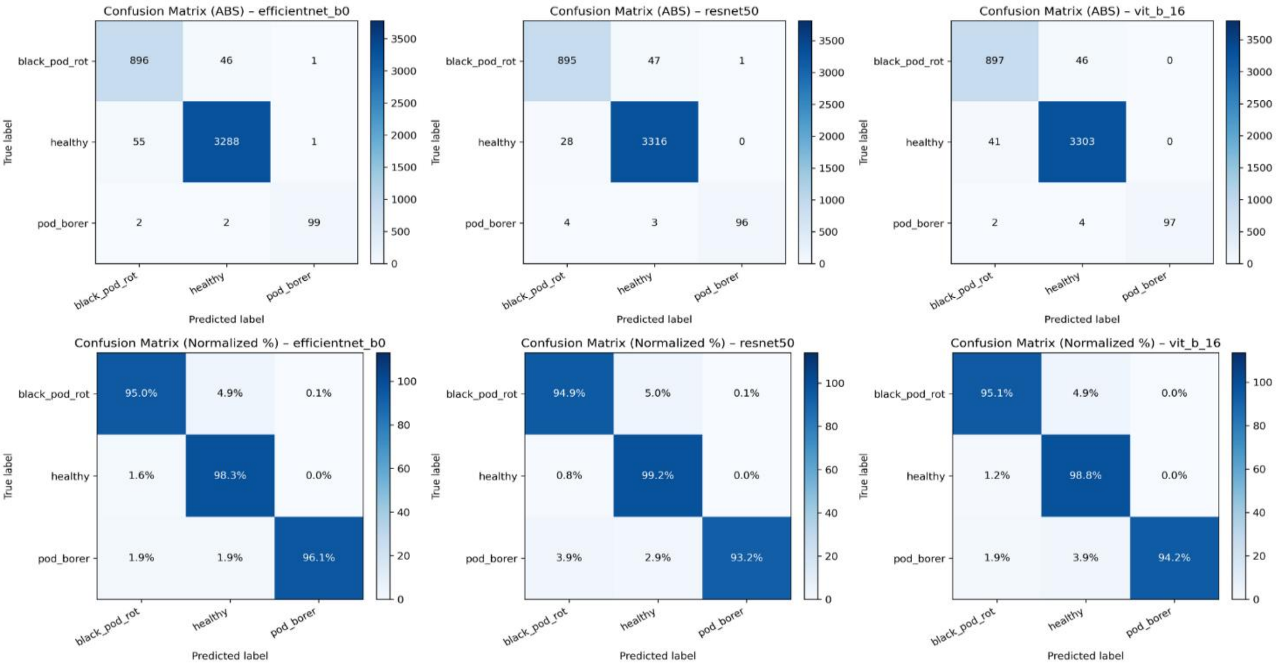

Figure 6. Confusion matrices by model

3.5 Inferential statistical analysis (ANOVA and post-hoc)

Statistical comparison of mean performance among the three architectures was performed using a repeated measures analysis of variance (RM-ANOVA) applied to the macro F1 metric, considering the fold as the subject and the model as the within-subject factor. The results showed no statistically significant differences between architectures (F = 0.278, p = 0.7645), suggesting homogeneous behavior in the average performance of the models. This finding was corroborated by paired post-hoc tests with Bonferroni correction, in which none of the pairwise comparisons reached statistical significance. In particular, the mean macro F1 difference between ViT-B/16 and EfficientNet-B0 was approximately +0.0031 with a 95% confidence interval and p ≈ 0.25, while the difference between ViT-B/16 and ResNet50 was +0.0014, also not significant. Taken together, these results indicate that, although ViT-B/16 had the highest mean, the observed discrepancies were small and statistically attributable to sampling variability, confirming the robustness of the three architectures evaluated (Table 3).

Table 3. Post-hoc tests by pair of models

Models	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

Practical Implications

The results obtained demonstrate that the three architectures evaluated offer competitive and stable performance for the automatic classification of cacao diseases, with similar macro F1 values and no statistically significant differences. This finding confirms that the model's effectiveness does not depend strictly on the architecture used, but rather on the coherence of the preprocessing pipeline, stratified validation, and overfit control. Consequently, the choice of reference model can be guided by operational and contextual implementation criteria: Vision Transformer (ViT-B/16), due to its slight average advantage and ability to model global spatial relationships; ResNet50, due to its maturity, broad support in production libraries, and behavior very similar to the former; and EfficientNet-B0, when computational efficiency and reducing the deployment footprint are prioritized.

These results support the findings of Ray et al. (2025) and Deepa et al. (2025), who highlight that deep learning models applied to agricultural diagnostics offer high levels of accuracy even with lightweight architectures, provided they are accompanied by consistent training strategies and adequately preprocessed data. They also confirm the trend observed by Lebrini & Ayerdi Gotor (2024) and Shafay et al. (2025), regarding the fact that the effectiveness of computer vision in the field of plant health depends more on methodological rigor than on the complexity of the model itself.

From an applied perspective, the results help fill a methodological gap identified in recent literature: the lack of systematic comparisons between contemporary architectures under controlled statistical frameworks. In this sense, the study provides reproducible evidence that guides the selection of models based on the balance between accuracy, stability, and feasibility of implementation in resource-constrained agricultural environments, a line of research highlighted by Bono et al. (2026) in the context of smart agriculture. Finally, future improvements should focus not so much on replacing architectures, but on optimizing complementary strategies such as class rebalancing, augmenting lesion-specific data, and adaptive adjustment of decision thresholds, following the recommendations of Song et al. (2025) on the need to integrate robust pipelines that maximize generalization in agricultural computer vision applications.

CONCLUSIONS

This study demonstrated that the evaluated deep learning architectures exhibit statistically equivalent performance in the automated detection of cacao pests and diseases, achieving macro F1 values above 0.96 with low standard deviations across folds. These results highlight the robustness of the implemented computer vision pipeline and the effectiveness of the preprocessing, validation, and regularization strategies employed. The absence of significant differences among models indicates that approach selection can be guided by operational criteria, considering the balance between accuracy, computational efficiency, and deployment scalability. Moreover, the findings confirm that system stability and generalization depend primarily on training quality and class imbalance handling rather than on the specific architecture used. Overall, these results contribute to the development of reproducible and scalable intelligent systems for phytosanitary monitoring, strengthening the integration of artificial intelligence into precision agriculture and its application in real-world production contexts.

FINANCING

The authors received no funding or sponsorship to conduct this study/article.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest related to the subject matter of this work.

AUTHORSHIP CONTRIBUTION

Conceptualization: Navarro-Cabrera, J. R. Data curation and formal analysis: Beraún-Barrantes, J. G.; Cárdenas-García, Á.; Lozano-Carranza, C. M. Research: Navarro-Cabrera, J. R.; Beraún-Barrantes, J. G. Methodology and project management: Navarro-Cabrera, J. R. Software: Navarro-Cabrera, J. R. Validation: Navarro-Cabrera, J. R.; Beraún-Barrantes, J. G. Visualization: Navarro-Cabrera, J. R. Writing – original draft: Navarro-Cabrera, J. R.; Beraún-Barrantes, J. G.; Cárdenas-García, Á.; Lozano-Carranza, C. M. Writing – revision and editing: Navarro-Cabrera, J. R.

REFERENCES

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/repiama.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—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-Ikhlas, 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