Salud, Ciencia y Tecnología - Serie de Conferencias. 2025; 4:1762
doi: 10.56294/sctconf20251762

ORIGINAL

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

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

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

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

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

Submitted: 12-07-2025 **Revised:** 09-09-2025 **Accepted:** 24-11-2025 **Published:** 25-11-2025

Editor: Dr. William Castillo-González

Corresponding author: Jorge Raúl Navarro-Cabrera

ABSTRACT

Introduction: deep learning (DL)-based computer vision has emerged as a promising tool for precision agriculture, particularly for detecting crop diseases and pests automatically. This study evaluated the comparative performance of three state-of-the-art DL architectures for automatic identification of cacao pests and diseases using image analysis.
Method: a reproducible pipeline was implemented, encompassing image preprocessing, stratified cross-validation, and inferential statistics through repeated-measures ANOVA. The dataset comprised 4,390 images divided into three highly unbalanced classes: Healthy, Black Pod Rot, and Pod Borer. The architectures—ResNet50, EfficientNet-B0, and ViT-B/16—were fully fine-tuned using the AdamW optimizer, early stopping, and a dynamic learning-rate scheduler.
Results: all models achieved mean macro-F1 scores above 0,96, with no statistically significant differences observed among them (F = 0,278, p = 0,7645). Training curves showed rapid convergence and inter-fold stability, indicating consistent generalization without overfitting.
Conclusions: performance outcomes suggest that the effectiveness of the detection system relies more on pipeline design and class-balance management than on the specific DL architecture used. The findings contribute to developing reproducible, efficient intelligent systems for cacao phytosanitary monitoring and support the integration of artificial intelligence into precision agriculture practices.

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

RESUMEN

Introducción: el aprendizaje profundo (deep learning, DL) aplicado a la visión por computadora se ha consolidado como una herramienta prometedora en la agricultura de precisión, especialmente para la detección automática de plagas y enfermedades en cultivos. El presente estudio evaluó el desempeño comparativo de tres arquitecturas avanzadas de DL para la identificación automática de plagas y enfermedades del cacao mediante análisis de imágenes.
Método: se implementó un pipeline reproducible que incluyó preprocesamiento de imágenes, validación cruzada estratificada y análisis estadístico inferencial mediante ANOVA de medidas repetidas. El conjunto

de datos comprendió 4 390 imágenes distribuidas en tres clases altamente desbalanceadas: Healthy, Black Pod Rot y Pod Borer. Las arquitecturas ResNet50, EfficientNet-B0 y ViT-B/16 fueron ajustadas completamente (fine-tuning total) y entrenadas con el optimizador AdamW, parada temprana y programador dinámico de tasa de aprendizaje.
Resultados: todos los modelos alcanzaron valores medios de F1 macro superiores a 0,96, sin diferencias estadísticamente significativas entre ellos (F = 0,278, p = 0,7645). Las curvas de entrenamiento mostraron una convergencia rápida y estabilidad entre pliegues, sin evidencias de sobreajuste.
Conclusiones: los resultados indican que la eficacia del sistema depende en mayor medida del diseño del pipeline y del control del desbalance de clases que del tipo de arquitectura empleada. Los hallazgos aportan a la implementación de sistemas inteligentes reproducibles y eficientes para el monitoreo fitosanitario del cacao, promoviendo la adopción de soluciones basadas en inteligencia artificial en la agricultura de precisión.

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

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 generating income for millions of small producers.[1,2,3,4] Its fermented and dried beans are the essential raw material for the chocolate industry and various confectionery products.[5,6] 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.[7,8,9] Among the most relevant pathologies are Black Pod Rot, caused by *Phytophthora* spp., and Pod Borer, both responsible for significant economic losses in tropical production areas.[10,11,12]

Traditionally, the detection of these conditions has been based on visual inspections carried out by farmers or specialists, a subjective procedure that depends on individual experience and morphological interpretation of symptoms.[13,14] This approach, in addition to being slow and costly, presents high variability in diagnostic accuracy, which delays the application of effective control measures and increases the spread of the disease. Consequently, the agricultural industry demands more reliable, rapid, and reproducible methods for the early identification of pests and diseases, which will optimize management processes and improve crop productivity.[15,16,17]

In this context, advances in computer vision and machine learning (ML) have opened up new opportunities for the automation of diagnostic tasks in precision agriculture.[18,19] These technologies allow large volumes of visual data to be analyzed using algorithms capable of recognizing complex patterns in images of leaves, fruits, or stems.[20,21] By combining optical hardware and digital image processing software, artificial vision systems (AVS) have established themselves as non-destructive, fast, and cost-effective tools for the characterization and classification of agricultural products, achieving levels of accuracy comparable to and even superior to those obtained by human experts.[22,23,24]

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.[25,26] These approaches allow relevant features to be automatically extracted from images, eliminating the need for manual attribute engineering and increasing the capacity for generalization in complex tasks.[27] Thanks to its multilevel architecture, DL has demonstrated outstanding performance in the detection, segmentation, and classification of visual patterns, establishing itself as the dominant paradigm in agricultural diagnosis, quality control, and automated crop monitoring applications.[28,29,30]

However, the current literature reveals significant limitations. Most studies focus on a single architecture or evaluate models without applying statistical methods that allow their performance to be compared with inferential significance. This lack of rigorous comparative analysis makes it difficult to clearly identify which architectures are most stable, accurate, and efficient in agricultural contexts with limited computational resources. In addition, most experiments lack cross-validation schemes or reproducible protocols, making it difficult to replicate results and adopt them in practice.

Given this gap, there is a need for systematic evaluations that compare contemporary deep learning approaches under a controlled experimental framework. These models present different trade-offs between accuracy, complexity, and computational efficiency, making it essential to analyze their performance in automatic classification tasks for agricultural pests and diseases. Evaluating their comparative behavior allows technical guidelines to be established for their implementation in intelligent monitoring and automated phytosanitary diagnosis systems, promoting solutions that combine high performance, stability, and operational viability in real production environments.[31]

This study proposes a deep learning-driven computer vision approach aimed at the early and automatic detection of pests and diseases in cocoa fruits. A reproducible methodological process was implemented that integrates 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 different phytosanitary conditions and evaluate the stability and accuracy of the models in automated classification scenarios. This work seeks to provide technical and practical evidence for the development of intelligent agricultural diagnostic systems, promoting sustainability and efficiency in global cocoa production.[32]

METHOD

The experimental development was structured in a reproducible computer vision pipeline designed for the automatic classification of diseases in cocoa fruits. The process comprised the following phases: (1) data exploration (EDA), (2) preprocessing, (3) training with stratified cross-validation, (4) evaluation using classification metrics, and (5) inferential statistical analysis to compare model performance. This methodological framework ensures the traceability of results and comparability between configurations, allowing the study to be replicated in different production contexts or with new datasets.

Data set and initial preparation

The image set called "Cacao Diseases" was used, obtained through Kaggle (folders by class), consisting of 4390 RGB images of cacao fruits labeled in three categories: Black Pod Rot (943), Pod Borer (103), and Healthy (3344). The images were organized under the ImageFolder scheme (one folder per class) for direct consumption by PyTorch. Before training, a basic EDA was performed: first, a count per class to identify imbalances, then sampling of image sizes to estimate resolution variability, and finally, a visual mosaic of examples per class. Any corrupt or unreadable images were discarded.

Experimental design and partitioning

To obtain robust estimates and enable statistical comparison between architectures, 5-fold stratified cross-validation was applied (same distribution per class in each fold). In each fold, each of the three models was trained and validated on the same partitions, which subsequently allowed the fold to be treated as a "subject" in the repeated measures analysis. A global seed (42) was set to ensure reproducibility.

Preprocessing and augmentations

In order to preserve color signals relevant to symptoms such as spots, mycelium, or discoloration, moderate magnification parameters and standard ImageNet normalization were used:

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 set to 224×224px to allow compatibility with the ImageNet standard and improve computational efficiency. The transformations in the augmenters were conservative, with the aim of not distorting fine features in the lesions.

Models evaluated

Three deep learning architectures widely used in recent studies of agro-phytopathological image classification were evaluated: ResNet50[33], EfficientNet-B0[34], and Vision Transformer (ViT-B/16)[35]. All networks were pre-

https://doi.org/10.56294/sctconf20251762 ISSN: 2953-4860

trained on the ImageNet-1K dataset, with the aim of leveraging general visual representations learned from millions of natural images. Subsequently, the output layers were reconfigured to adapt them to the specific domain of the problem. In the case of ResNet50, the final layer was replaced by a fully connected layer with three neurons corresponding to the classes of interest. In EfficientNet-B0, the last layer of the classifier was replaced by a three-class output, while in ViT-B/16, the original classification component was replaced by a linear layer with the same output dimensionality.

In all three models, a complete adjustment of the parameters (total fine-tuning) was performed, instead of freezing the pre-trained convolutional or attention layers. This methodological decision responds to the substantial difference between the source domain, represented by ImageNet, based on natural objects and general scenes, and the target domain of the present study, focused on images of cocoa fruits affected by lesions and pests. Comprehensive weight adjustment allows architectures to learn more specific discriminative representations of the phytopathological context, optimizing the model's sensitivity to subtle visual patterns and textures associated with different fruit health states.

Training configuration

Model training was implemented using the PyTorch framework, applying a uniform set of hyperparameters to ensure comparability between architectures. The AdamW[36] optimizer was used with an initial learning rate of 1×10^{-4} and a weight decay parameter of 1×10^{-4}, configurations that balance gradient stability and model regularization during learning. The selected loss function was Cross-Entropy Loss, suitable for multi-class classification problems. The batch size was set at 32 images per iteration, and the training process was run 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, which reduces the learning rate value by 50 % when the validation loss shows no significant improvement. Likewise, an early stopping mechanism was implemented with a patience of three epochs, selecting as the optimal point the checkpoint corresponding to the highest macro F1 value achieved in validation. This set of strategies made it possible to control overfitting, stabilize convergence, and optimize the use of computational resources, which is suitable for both environments with hardware limitations and inference scenarios on edge devices.

Fold evaluation protocol

At the end of each epoch, the fold validation set was evaluated and, once the stopping criterion was activated, the best model state was loaded to compute the final fold metrics: Accuracy, Macro F1, Macro Precision, and Macro Recall. A classification report by class (precision/recall/F1) and the confusion matrix were also generated.

The confusion matrices per fold were stored and then aggregated by model in two ways: by sum of absolute counts and by a normalized version per row (recall per class), in order to analyze systematic confusion patterns between classes without a single fold dominating the interpretation.

Statistical comparison between models

To compare the average performance of the three architectures, a statistical design of repeated measures was applied, considering the five folds generated in the cross-validation. First, a long-format table was constructed that included as factors the fold (treated as a subject), the model (considered as an intra-subject factor with three levels), and the F1 macro response metric. Subsequently, a repeated measures analysis of variance (ANOVA-RM) was performed using the model as an intra-subject effect and the fold as the unit of observation, establishing a significance level of $\alpha = 0{,}05$. In cases where statistical significance was detected, a post-hoc analysis was performed using paired t-tests between each pair of models, applying the Bonferroni correction to control for cumulative type I error. This approach allowed for a robust evaluation of the differences attributable to architecture, controlling for the variability introduced by the validation partitions and avoiding conclusions dependent on a single training-validation split.

RESULTS

Data exploration (EDA)

After initial exploration of the dataset, 4,390 images were counted, distributed across three classes with considerable imbalance: the Healthy category contained 3,344 images (76,2 %), while Black Pod Rot had 943 (21,5 %) and Pod Borer only 103 (2,35 %), representing an approximate ratio of 32,5 times between the majority and minority classes. Geometric analysis performed on a sample of 300 images indicated that all captures have 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 the proportions per class in each fold, as well as the prioritization of the macro F1 metric in the selection of checkpoints in order to mitigate the bias induced by class imbalance. Likewise, it was decided to apply moderate data augmentations during training, seeking to improve the model's generalization ability without compromising the integrity of the visual signals associated with the lesions (figure 1).

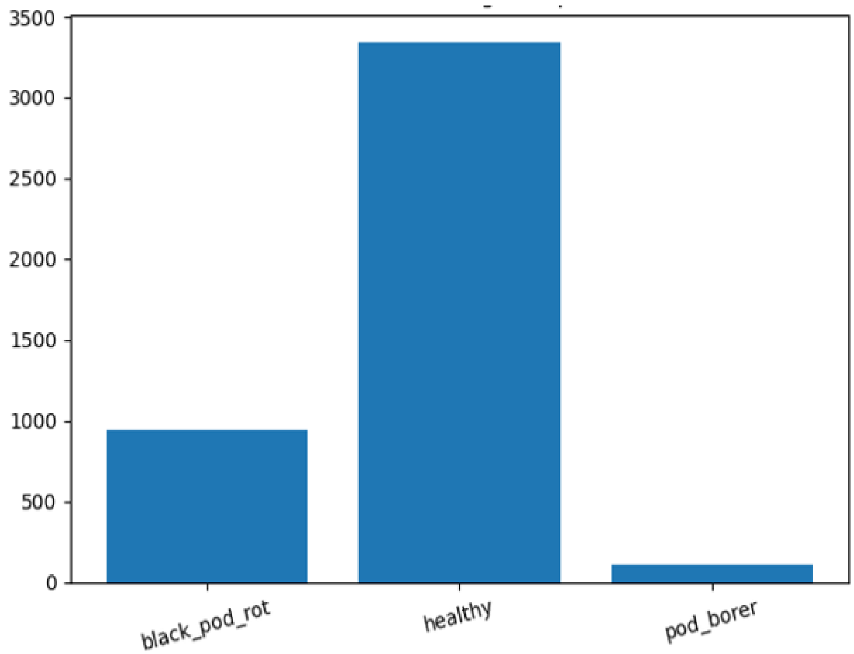

Figure 1. Class distribution graph

Overall model performance

Analysis of the overall performance of the models showed highly competitive performance among the three architectures evaluated. The Vision Transformer (ViT-B/16) model achieved the highest macro F1 mean, 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 an average macro F1 of 0,9683, followed by EfficientNet-B0, which scored 0,9666. The differences between the three models were minimal, with an average separation of approximately 0,3 percentage points between the best and worst performers in the macro F1 metric. This closely grouped behavior reflects the robustness of the 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	Accuracy Average	Dev. Accuracy	F1 mean	F1 Deviation	Prec. Average	Prec. Dev.	Recall Average	Recall Deviation
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

Statistical comparison (ANOVA and post-hoc)

The statistical comparison of the average performance between the three architectures was performed using a repeated measures analysis of variance (ANOVA-RM) applied to the F1 macro metric, considering the fold as the subject and the model as the intra-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 contrasts reached statistical significance. In particular, the mean difference in macro F1 between ViT-B/16 and EfficientNet-B0 was approximately +0,0031 with a 95 % confidence interval and $p \approx 0,25$, 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 discrepancies observed were small and statistically attributable to sampling variability, confirming the robustness of the three architectures evaluated (table 2).

https://doi.org/10.56294/sctconf20251762 ISSN: 2953-4860

Table 2. Post-hoc tests by model pair		
Models	**t**	**p_raw**
ResNet50 vs EfficientNet B0	0,297661326	0,780781239
ResNet50 vs ViT-B/16	-0,370845959	0,729557
EfficientNet B0 vs ViT-B/16	-1,347132941	0,249193532

The evolution of the macro F1 during the training and validation process showed a stable convergence pattern in the three models, with no signs of overfitting and minimal fluctuations between folds. Figures 2, 3, and 4 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.

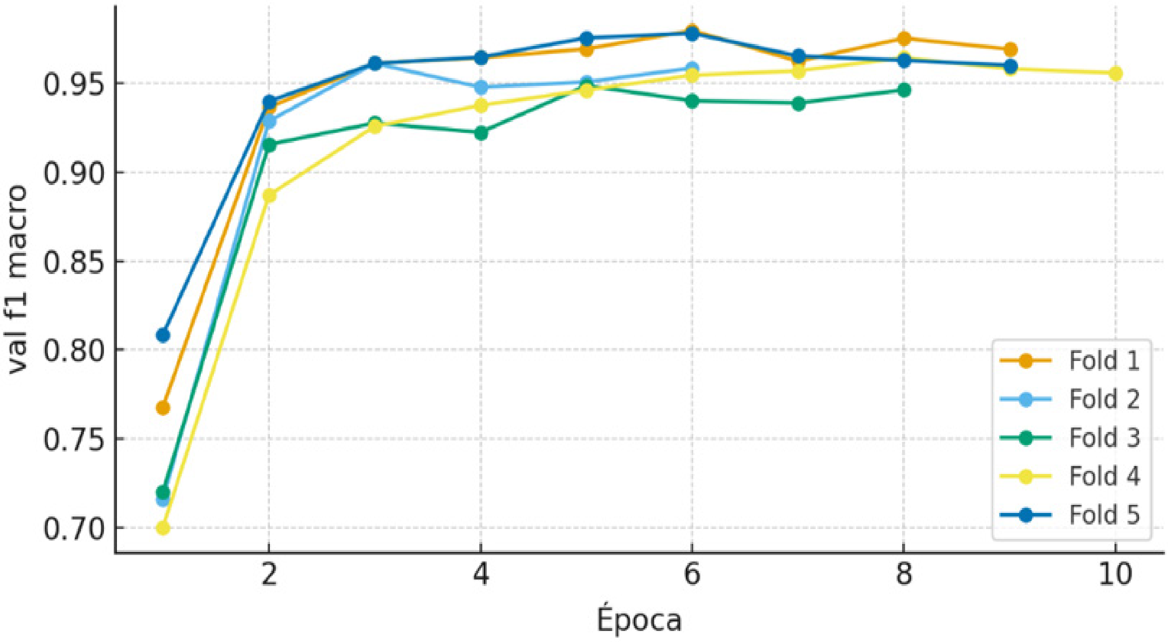

Figure 2. Evolution of F1 Macro (validation) per fold in the EfficientNetB0 model

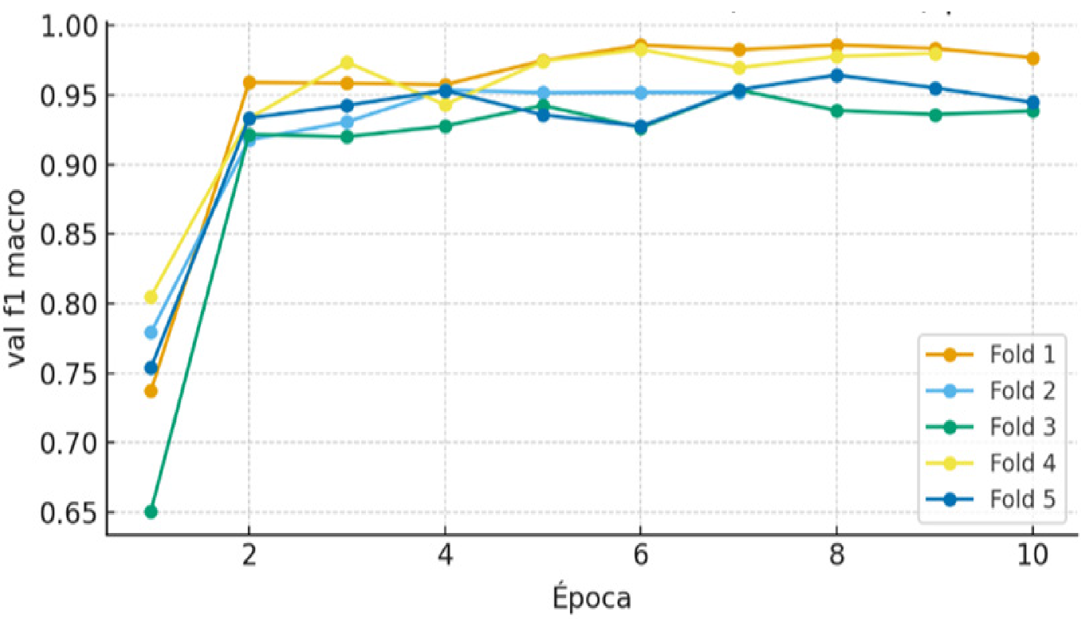

Figure 3. Evolution of F1 Macro (validation) per fold in ResNet50 model

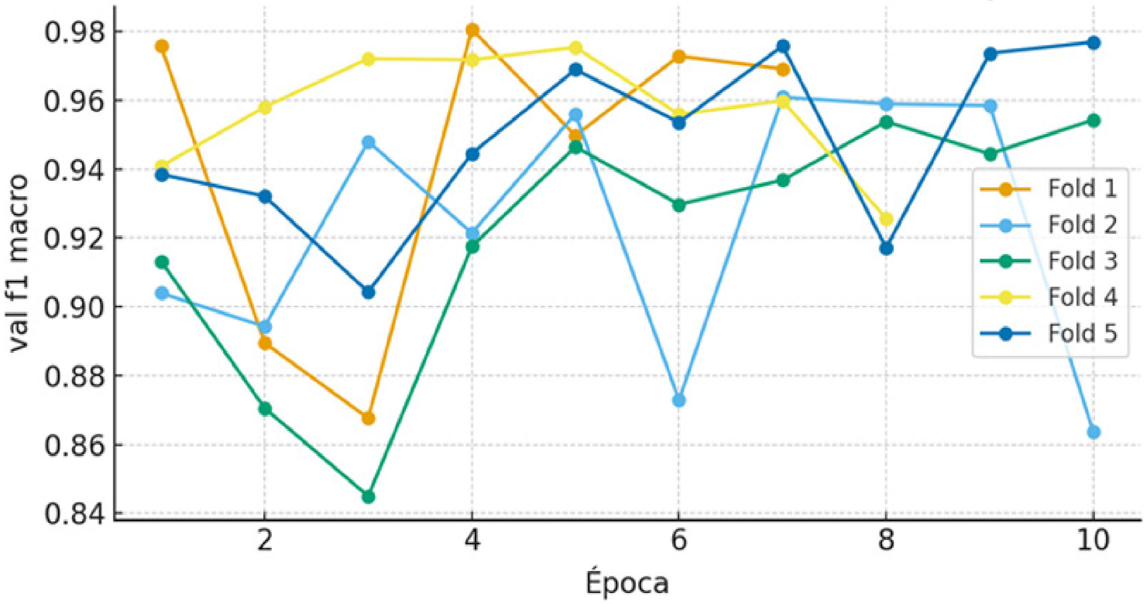

Figure 4. Evolution of Macro F1 (validation) per fold in the ViT-B/16 model

Consistency between folds and training dynamics

The consistency analysis between folds showed low variability in all experimental configurations, with standard deviations of the F1 macro metric close to one hundredth, reflecting high stability in the training and validation process. This behavior suggests that the decisions made during image preprocessing and the use of stratified cross-validation contributed to the robustness of the procedure, avoiding the influence of a dominant fold that could skew the overall averages. The detailed results for each architecture in the five folds are presented in table 3, which shows the consistency of the accuracy, precision, recall, and F1 metrics, evidencing balanced performance between models and partitions.

Likewise, the training and validation curves showed rapid convergence, reaching stability in less than ten epochs. The use of the learning rate scheduler favored the stabilization of loss in 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 3. 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
5	ViT-B/16	0,9841	0,9769	0,9830	0,9713

The confusion matrices aggregated by model, obtained from the sum of the five folds, together with their row-normalized versions, showed balanced behavior between classes, with no evidence of dominant systematic confusion patterns. Row normalization allowed for a detailed analysis of recall by class, showing slight oscillations between architectures that are consistent with the narrow range of macro F1 values previously reported. These variations do not alter the overall conclusion of the study: the three models show a consistent ability to correctly discriminate between the Black Pod Rot, Pod Borer, and Healthy categories, maintaining homogeneous performance across the evaluated dataset (figure 5).

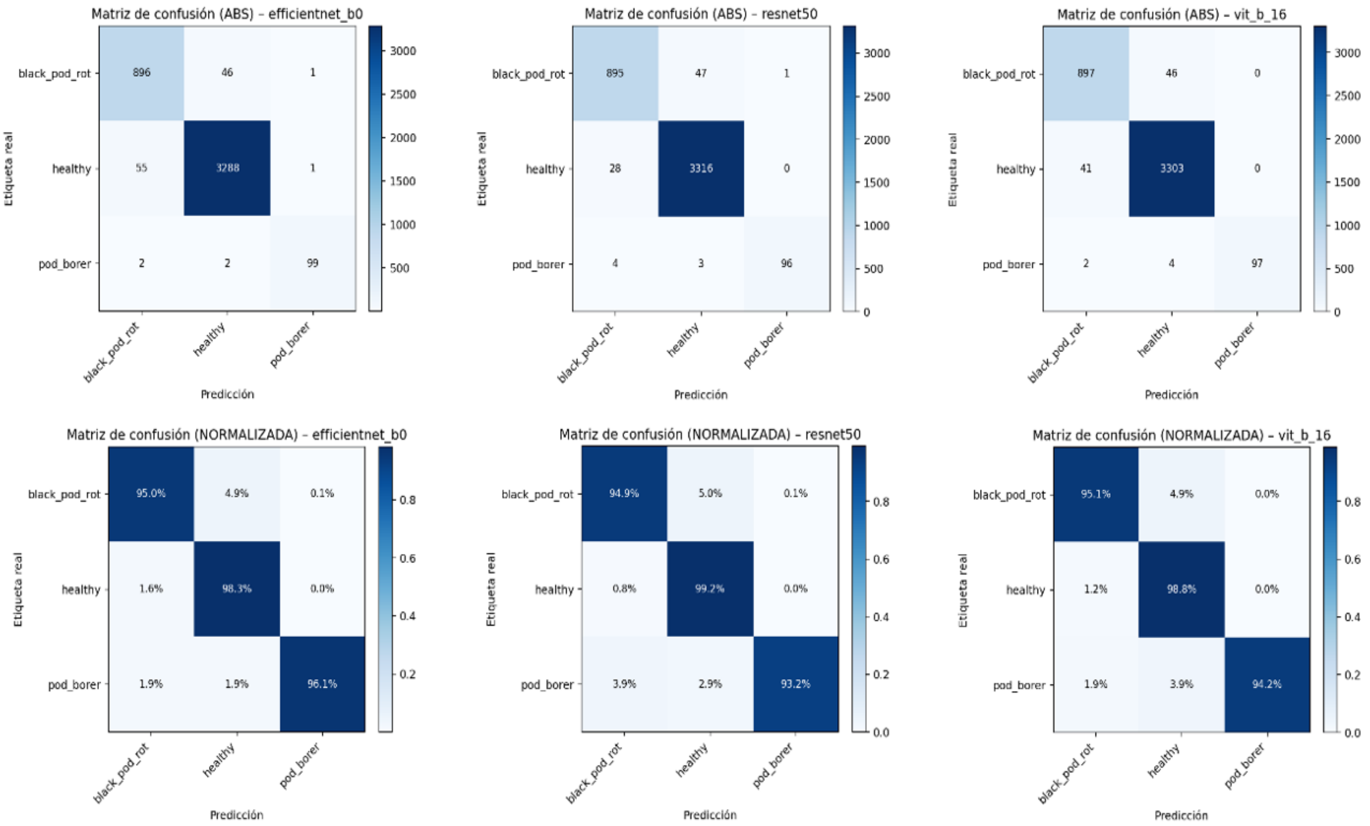

Figure 5. Confusion matrices by model

DISCUSSION

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

These results support the findings of Ray et al.[25] and Deepa et al.[28], who highlight that deep learning models applied to agricultural diagnosis 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 and Ayerdi Gotor[21] and Shafay et al.[29], namely that the effectiveness of computer vision in the phytosanitary field depends more on methodological rigor than on the complexity of the model itself.

From an applied perspective, the results contribute to filling 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.[31] 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, increases in specific data by injury type, and adaptive adjustment of decision thresholds, following the recommendations of Song et al.[22] on the need to integrate robust pipelines that maximize generalization in agricultural computer vision applications.

CONCLUSIONS

The study demonstrated that the deep learning architectures evaluated show statistically equivalent performance in the automatic detection of cocoa pests and diseases, achieving macro F1 values above 0,96 and reduced standard deviations between folds. These results demonstrate the robustness of the implemented computer vision pipeline and the effectiveness of the preprocessing, validation, and regularization strategies employed. The absence of significant differences between models indicates that the selection of the approach can be guided by operational criteria, considering the balance between accuracy, computational efficiency, and scalability of deployment. Likewise, it is confirmed that the stability and generalization of the system depend mainly on the quality of training and the handling of class imbalance, rather than on the type of architecture used. Taken together, the findings contribute to the development of reproducible and scalable intelligent systems for phytosanitary monitoring, strengthening the integration of artificial intelligence in precision agriculture and its application in real production contexts.

REFERENCES

1. Charry A, Perea C, Ramírez K, Zambrano G, Yovera F, Santos A, et al. The bittersweet economics of different cacao production systems in Colombia, Ecuador and Peru. Agric Syst. 2025;224:104235. https://doi.org/10.1016/j.agsy.2024.104235.

2. Vinci G, Ruggeri M, Gobbi L, Savastano M. Social Life Cycle Assessment of Cocoa Production: Evidence from Ivory Coast and Ghana. Resources. 2024;13:141. https://doi.org/10.3390/resources13100141.

3. Angulo Rincón SO, Solarte Solarte CM. Green innovation as a sustainable alternative to strengthen cocoa cultivation. Environ Res Ecotoxicity. 2026;5:300. https://doi.org/10.56294/ere2026300.

4. Angulo Rincón SO, Solarte Solarte CM. Green innovation analysis for cocoa farmers in the municipality of Roberto Payán. Environ Res Ecotoxicity. 2024;3:120. https://doi.org/10.56294/ere2024120.

5. Quintero I, Ceccaldi A, Martínez H, Santander M, Rodríguez J, Escobar S. Dry cacao pulp in chocolate bars: A sustainable, nutrient-rich sweetener with enhanced sensory quality through refractance windows drying. Appl Food Res. 2025;5:100700. https://doi.org/10.1016/j.afres.2025.100700.

6. Paparella A, Schirone M, López CC. The Health Impact of Cocoa from Cultivation to the Formation of Biogenic Amines: An Updated Review. Foods. 2025;14:255. https://doi.org/10.3390/foods14020255.

7. Delgado-Ospina J, Molina-Hernández JB, Chaves-López C, Romanazzi G, Paparella A. The Role of Fungi in the Cocoa Production Chain and the Challenge of Climate Change. J Fungi. 2021;7:202. https://doi.org/10.3390/jof7030202.

8. Zahlul Ikhsan, Muhammad Al-Ikhlas, Yaherwandi, Hasmiandy Hamid, Aulia Oktavia, Dandy Ahamefula Osibe. Level of Pest Infestation on Cocoa (Theobroma cacao L.) Variety BL-50 in Tanah Datar Regency, West Sumatera Province, Indonesia. Andalasian Int J Entomol. 2024;2:38–47. https://doi.org/10.25077/aijent.2.1.38-47.2024.

9. Cilas C, Bastide P. Challenges to Cocoa Production in the Face of Climate Change and the Spread of Pests and Diseases. Agronomy. 2020;10:1232. https://doi.org/10.3390/agronomy10091232.

10. Puig AS, Irish B, Ayala-Silva T, Wurzel S, Gutierrez O. Effect of Cacao Black Pod Rot Screening Method on Disease Reaction Determination. 1st Int Online Conf Agric Agric Sci Technol. Basel Switzerland: MDPI; 2022. p. 71. https://www.mdpi.com/2673-4583/10/1/71.

11. Schmidt JE, Puig AS, DuVal AE, Pfeufer EE. Phyllosphere microbial diversity and specific taxa mediate within-cultivar resistance to Phytophthora palmivora in cacao. mSphere. 2023;8. https://doi.org/10.1128/msphere.00013-23.

12. Magfirah N, Sartiami D, Niogret J, Ekayanti A. Biological parameter of cocoa pod borer Conopomorpha cramerella (Lepidoptera: Gracillariidae) under laboratory conditions. IOP Conf Ser Earth Environ Sci. 2025;1494:012020. https://doi.org/10.1088/1755-1315/1494/1/012020.

13. Miyittah MK, Kosivi RK, Tulashie SK, Addi MN, Tawiah JY. The need for alternative pest management methods to mitigate risks among cocoa farmers in the Volta region, Ghana. Heliyon. 2022;8:e12591. https://doi.org/10.1016/j.heliyon.2022.e12591.

14. Polania Bello V. Diagnosis of the agronomic and phytosanitary management practices of cocoa producers in San José del Fragua: the case of the Cerafín Garcia property. Environ Res Ecotoxicity. 2023;2:55. https://doi.org/10.56294/ere202355.

15. Wu K, Ji Z, Wang H, Shao X, Li H, Zhang W, et al. A Comprehensive Review of AI Methods in Agri-Food Engineering: Applications, Challenges, and Future Directions. Electronics. 2025;14:3994. https://doi.org/10.3390/electronics14203994.

16. Raj M, Prahadeeswaran M. Revolutionizing agriculture: a review of smart farming technologies for a sustainable future. Discov Appl Sci. 2025;7:937. https://doi.org/10.1007/s42452-025-07561-6.

17. Tirado J. Big data and artificial intelligence for innovation in management in the agricultural industry in Extremadura. EthAIca. 2022;1:38. https://doi.org/10.56294/ai202238.

18. Waqas M, Naseem A, Humphries UW, Hlaing PT, Dechpichai P, Wangwongchai A. Applications of machine learning and deep learning in agriculture: A comprehensive review. Green Technol Sustain. 2025;3:100199. https://doi.org/10.1016/j.grets.2025.100199.

19. Taha MF, Mao H, Zhang Z, Elmasry G, Awad MA, Abdalla A, et al. Emerging Technologies for Precision Crop Management Towards Agriculture 5.0: A Comprehensive Overview. Agriculture. 2025;15:582. https://doi.org/10.3390/agriculture15060582.

20. Injante R, Rios-Trigoso G, Ramírez-Shupingahua S, Tejada Shupingahua K. Procesamiento de imágenes para la detección de la madurez del fruto: una revisión sistemática. Rev Peru Ing Arquit y Medio Ambient. 2025;2. https://doi.org/10.37711/repiama.2025.2.1.3.

21. Lebrini Y, Ayerdi Gotor A. Crops Disease Detection, from Leaves to Field: What We Can Expect from Artificial Intelligence. Agronomy. 2024;14:2719. https://doi.org/10.3390/agronomy14112719.

22. Song X, Yan L, Liu S, Gao T, Han L, Jiang X, et al. Agricultural Image Processing: Challenges, Advances, and Future Trends. Appl Sci. 2025;15:9206. https://doi.org/10.3390/app15169206.

23. Anjali, Jena A, Bamola A, Mishra S, Jain I, Pathak N, et al. State-of-the-art non-destructive approaches for maturity index determination in fruits and vegetables: principles, applications, and future directions. Food Prod Process Nutr. 2024;6:56. https://doi.org/10.1186/s43014-023-00205-5.

24. Uriarte del **Águila** C. Application of data mining as a strategic tool in knowledge management. Diginomics. 2024;3:123. https://doi.org/10.56294/digi2024123.

25. Ray RK, Chakravarty S, Dash S, Ghosh A, Mohanty SN, Reddy Chirra VR, et al. Precision pest management in agriculture using Inception V3 and EfficientNet B4: A deep learning approach for crop protection. Inf Process Agric. 2025. https://doi.org/10.1016/j.inpa.2025.09.005.

26. Villalobos-Culqui C, Valles-Coral MA, García-Rivas-Plata C, Tuesta-Hidalgo OA. Modelo de visión artificial basada en redes neuronales convolucionales para identificación de mazorca negra en plantaciones de cacao. Rev Científica Sist e Informática. 2025;5:e678. https://doi.org/10.51252/rcsi.v5i1.678.

27. Mall PK, Singh PK, Srivastav S, Narayan V, Paprzycki M, Jaworska T, et al. A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. Healthc Anal. 2023;4:100216. https://doi.org/10.1016/j.health.2023.100216.

28. Deepa R, Varun P, Shoba LK, Swathy R, Prabhu B. Advanced Image Processing Techniques for Precision Agriculture. Innov Dev Unmanned Aer Veh. IGI Global Scientific Publishing; 2025. p. 227–244. https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/979-8-3693-8462-6.ch010.

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

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

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

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

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

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

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

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

FINANCING
The authors did not receive funding for the development of this research.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTION
Conceptualization: Jorge Raúl Navarro-Cabrera.
Data curation: Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes.
Formal analysis: Jorge Raúl Navarro-Cabrera, Ángel Cárdenas-García.
Research: Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes, Ángel Cárdenas-García, Carlos Mauricio Lozano-Carranza.
Methodology: Jorge Raúl Navarro-Cabrera.
Software: Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes.
Validation: Jorge Raúl Navarro-Cabrera, Carlos Mauricio Lozano-Carranza.
Visualization: Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes, Ángel Cárdenas-García.
Writing – original draft: Jorge Raúl Navarro-Cabrera, José Guillermo Beraún-Barrantes, Carlos Mauricio Lozano-Carranza.
Writing – revision and editing: Jorge Raúl Navarro-Cabrera, Ángel Cárdenas-García.