UNIVERSIDAD NACIONAL DE SAN MARTIN

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

Original article
Artículo original
Jan-Jun, 2026

Automatic detection of coffee leaf diseases through pattern recognition

Detección automática de enfermedades foliares del cafeto mediante reconocimiento de patrones

Santa-María, John C.[1,2*]

Rodríguez, Ciro[1]

[1]Universidad Nacional Mayor de San Marcos, Lima, Perú
[2]Universidad Nacional de San Martín, Tarapoto, Perú

Received: 15 May 2025 | **Accepted:** 22 Dec. 2025 | **Published:** 20 Jan. 2026

Corresponding author*: john.santamaria@unmsm.edu.pe

How to cite this article: Santa-María, J. C. & Rodríguez, C. (2026). Automatic detection of coffee leaf diseases through pattern recognition. *Revista Científica de Sistemas e Informática*, *6*(1), e1349. https://doi.org/10.51252/rcsi.v6i1.1349

ABSTRACT

Early detection of coffee leaf diseases is essential to reduce production losses; however, visual field diagnosis presents limitations associated with subjectivity and environmental variability. The objective of this study was to design and evaluate a hybrid pattern-recognition model to classify healthy coffee leaves and those affected by coffee leaf rust and brown eye spot using images captured under real field conditions in Saposoa (San Martín, Peru). A proprietary dataset of 1500 images validated by a specialist (500 per class) was used and expanded through controlled data augmentation to 6000 balanced images. ResNet18 was employed as a feature extractor using transfer learning, and three supervised classifiers were compared: SVM, Random Forest, and XGBoost. Model performance was evaluated using 10-fold stratified cross-validation and an independent test set (20%). The ResNet18 + SVM model achieved the best performance, with an accuracy of 0.9742, a macro F1-score of 0.9730, and a macro-AUC of 0.9968, outperforming Random Forest (accuracy = 0.9367) and XGBoost (accuracy = 0.9583). Inferential analysis using ANOVA and Tukey's HSD test confirmed statistically significant differences among models ($p < 0.001$). The results demonstrate the robustness and feasibility of the proposed approach to support coffee phytosanitary diagnosis under real-world field conditions.

Keywords: computer vision; deep learning; pattern recognition; phytosanitary diagnosis; transfer learning

RESUMEN

La detección temprana de enfermedades foliares del cafeto es clave para reducir pérdidas productivas; sin embargo, el diagnóstico visual en campo presenta limitaciones asociadas a la subjetividad y la variabilidad ambiental. El objetivo de este estudio fue diseñar y evaluar un modelo híbrido de reconocimiento de patrones para clasificar hojas de cafeto sanas, con roya y con ojo de gallo a partir de imágenes capturadas en condiciones reales en Saposoa (San Martín, Perú). Se empleó un conjunto de datos propio de 1500 imágenes validadas por especialista (500 por clase), ampliado mediante aumento de datos controlado hasta 6000 imágenes balanceadas. ResNet18 fue utilizado como extractor de características por transferencia de aprendizaje y se compararon tres clasificadores supervisados: SVM, Random Forest y XGBoost. La evaluación se realizó mediante validación cruzada estratificada de 10 pliegues y un conjunto de prueba independiente (20%). El modelo ResNet18 + SVM obtuvo el mejor desempeño, con una accuracy de 0.9742, F1-macro de 0.9730 y AUC-macro de 0.9968, superando a Random Forest (accuracy = 0.9367) y XGBoost (accuracy = 0.9583). El análisis inferencial mediante ANOVA y la prueba de Tukey HSD confirmó diferencias estadísticamente significativas entre modelos ($p < 0.001$). Los resultados evidencian la robustez y viabilidad del enfoque propuesto para apoyar el diagnóstico fitosanitario del cafeto en condiciones reales de campo.

Palabras clave: aprendizaje profundo; diagnóstico fitosanitario; reconocimiento de patrones; transferencia de aprendizaje; visión artificial.

1. INTRODUCTION

The early and accurate detection of coffee leaf diseases constitutes a critical factor for the productive and economic sustainability of smallholder farmers in Latin America. Among the most severe pathologies are coffee leaf rust (*Hemileia vastatrix*) and American leaf spot (*Mycena citricolor*), which are responsible for significant losses in yield and quality, as well as direct impacts on food security and the financial stability of coffee growers (Avelino et al., 2015; Poma-Angamarca et al., 2024).

Despite their agronomic relevance, the diagnosis of these diseases in the field continues to rely predominantly on visual inspection performed by farmers or technicians. This method presents significant limitations associated with evaluator subjectivity, variability in symptom manifestation, and the influence of environmental conditions such as lighting, humidity, and the plant's phenological stage, which increases the likelihood of diagnostic errors (Abade et al., 2020; Abdullah et al., 2023).

Inaccurate diagnoses lead to inadequate agronomic decisions, such as delayed or unnecessary fungicide applications, increasing production costs and reducing the effectiveness of phytosanitary management. This issue is particularly critical for smallholder farmers, who have limited resources and restricted access to specialized technical assistance, thereby amplifying the economic impact of these diseases (Julca-Otiniano et al., 2024; Saavedra-Ramírez, 2023).

In recent years, computer vision and deep learning have emerged as promising tools for the automatic detection of plant diseases. Numerous studies have shown that convolutional neural networks and transfer learning approaches enable the classification of foliar diseases with high levels of accuracy, including coffee plant pathologies such as leaf rust and cercospora (Archana & Jeevaraj, 2024; Mansouri et al., 2024; Martinez et al., 2022).

Additionally, hybrid models have been proposed that combine automatic feature extraction through deep learning with traditional supervised classifiers, such as SVM or Random Forest, achieving improvements in robustness and performance under variations in illumination and visual noise (Abuhayi & Mossa, 2023; Ayikpa et al., 2022). Nevertheless, challenges related to model generalization and reliability in real-world field scenarios persist.

A recurring limitation in the literature is that many models are trained using international datasets or images captured under controlled conditions, with isolated leaves and uniform backgrounds. These conditions differ significantly from real cultivation environments, where images exhibit shadows, occlusions, multiple infection stages, and high environmental variability, which adversely affects the performance of automatic diagnosis (Chavarro et al., 2023).

In the Peruvian context, and particularly in the San Martín region, coffee farming represents a fundamental economic and social pillar. Recent studies report the presence of new races of *Hemileia vastatrix* and a high incidence of American leaf spot in humid agroforestry systems, increasing the complexity of visual diagnosis and the need for technological tools adapted to local conditions (Ehrenbergerová et al., 2018; Julca-Otiniano et al., 2024).

Despite technological advances, a relevant scientific gap has been identified: the absence of robust automatic models trained on real images from Peruvian coffee plantations that enable the differentiated identification of coffee leaf rust and American leaf spot under conditions specific to the San Martín region. This gap limits the effective adoption of intelligent systems for local phytosanitary management.

In this context, the purpose of the present research is to design and evaluate a hybrid automatic analysis model based on pattern recognition, integrating computer vision and deep learning techniques to identify the presence of coffee leaf rust and American leaf spot in coffee leaves from images captured under field conditions. The model is trained using a proprietary dataset representative of the agroecological conditions of San Martín, with the aim of providing an objective, reproducible, and applicable diagnosis that complements traditional visual assessment and strengthens decision-making in integrated disease management.

2. MATERIALS AND METHODS

This research was conducted as an applied study, aimed at the design and evaluation of an automatic pattern recognition model for the classification of coffee leaf diseases using digital images captured under real field conditions. The level of the research was explanatory, as it analyzed how feature extraction through transfer learning, combined with supervised classifiers, enabled discrimination among healthy leaves and those affected by coffee leaf rust and American leaf spot. The research design was non-experimental and cross-sectional, since the phytosanitary condition of the leaves was not manipulated and the data were collected over a defined period. In addition, the study incorporated a technological approach by proposing and implementing a computational model with practical applicability in the phytosanitary diagnosis of coffee crops.

The materials used to carry out this research were divided into software and hardware, which are described in Table 1.

Table 1. Software and hardware employed

Component	Details
Software	Google Colab (training and evaluation), Google Drive (storage), Python (pipeline scripts), OpenCV and scikit-image (preprocessing), torch and torchvision (CNN feature extractor), scikit-learn (SVM/Random Forest and evaluation metrics), XGBoost (boosting classifier), pandas and matplotlib (analysis and visualization).
Hardware	Samsung Galaxy A54 5G mobile phone (field image acquisition).

Figure 1 illustrates the general methodology, structured as a computer vision system to classify coffee leaves into three phytosanitary classes. The methodology followed the following stages: (1) study area and design, (2) data collection and quality control, (3) dataset construction, (4) hybrid model development, and (5) evaluation and statistical comparison.

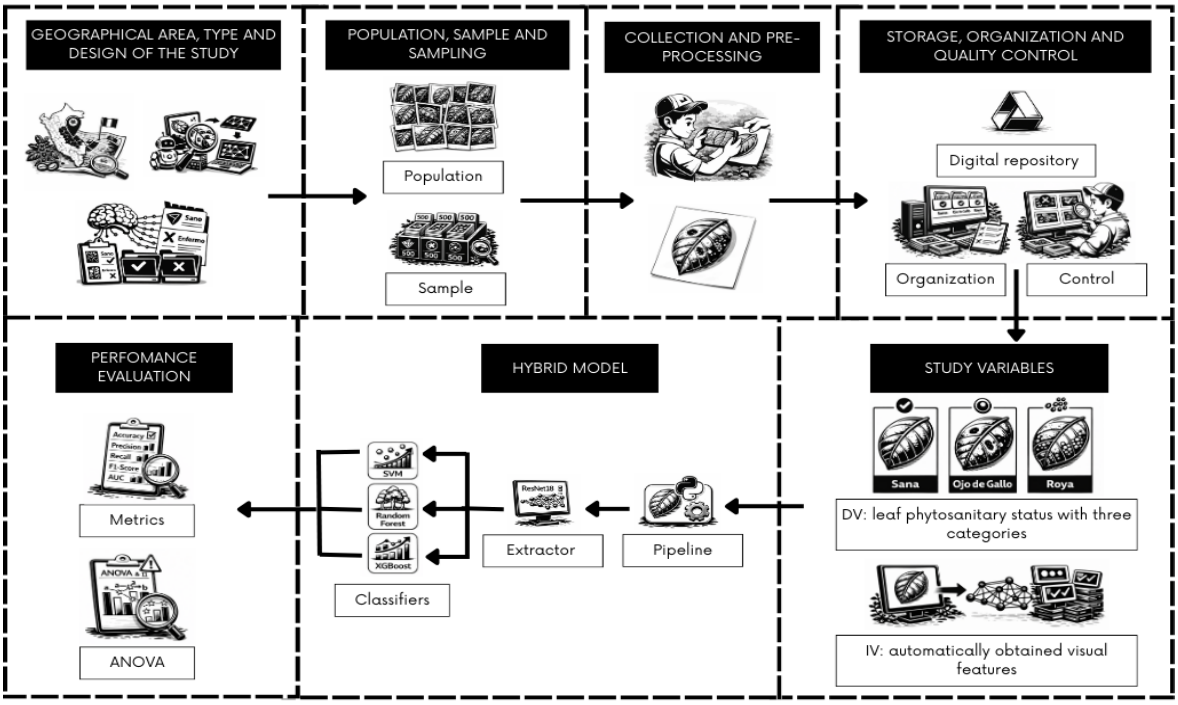

Figure 1. Methodological architecture of the hybrid pattern recognition model for the classification of coffee leaf diseases

Phase 1: Geographic area, type, and study design

The study was conducted in coffee plantations located in the district of Saposoa, San Martín region, Peru, under real field conditions. The research followed a technological–applied approach aimed at the design and evaluation of a pattern recognition model. The research level was explanatory, as it examined how feature extraction through transfer learning, combined with supervised classifiers, enabled the discrimination of disease classes. The design was non-experimental and cross-sectional, since the phytosanitary condition was not manipulated and image acquisition was carried out during a defined period.

Phase 2: Population, sample, and sampling

The population consisted of healthy and diseased coffee leaves present in plantations in Saposoa, recorded through digital images and with an initial diagnosis provided by plant health specialists. A sample of 1,500 images was used, distributed across three strata: 500 healthy leaves, 500 affected by American leaf spot, and 500 affected by coffee leaf rust. Selection was carried out using non-probabilistic convenience sampling, with stratification by phytosanitary condition; within each stratum, images with visual diversity and adequate technical quality were prioritized.

Inclusion criteria: (i) origin in Saposoa; (ii) capture under a uniform protocol (distance, lighting, and focus); (iii) classification confirmed by a specialist (healthy, American leaf spot, coffee leaf rust); (iv) sufficient quality (good resolution, correct focus, and absence of excessive shadows).

Exclusion criteria: (i) blurred, incomplete, or over-/underexposed images; (ii) leaves occluded by external objects; (iii) absence of specialist confirmation; (iv) intermediate stages without clear disease differentiation.

Phase 3: Image acquisition and preprocessing

Image acquisition was carried out directly in the field with the support of an agronomist, who evaluated each leaf prior to photographic recording. Field visits were conducted approximately between 10:00 a.m. and 1:00 p.m., a period with relatively uniform natural illumination. For image capture, each leaf was placed on a white background and photographed at a constant distance, ensuring full framing and adequate sharpness of edges and lesions. Images were stored in JPG format, preserving the native resolution of the device and avoiding additional compression processes.

Prior to model training, the images underwent a standardized preprocessing procedure. All samples were resized to 224×224 pixels, in accordance with the input requirements of the ResNet18 architecture, and were normalized using the mean and standard deviation parameters of the ImageNet dataset, in order to ensure compatibility with the pretrained model and numerical stability during training.

In addition, a controlled offline data augmentation strategy was applied to increase the visual variability of the dataset and improve the generalization capability of the models under real field conditions. For each original image, exactly three augmented images were generated using complementary transformations: (i) geometric augmentations (random rotations of ±50°, scaling and translation, with horizontal/vertical flips), (ii) color and illumination variations in the HSV color space (adjustments of hue, saturation, and brightness), and (iii) random cropping with resizing and the addition of Gaussian noise. Augmentation was performed offline to maintain traceability between original and augmented images, prevent information leakage during validation, and ensure a homogeneous balance among classes.

Phase 4: Storage, organization, and quality control

The images were transferred to a digital repository and organized into three directories: healthy, American leaf spot, and coffee leaf rust, according to the specialist's diagnosis. Quality control was applied at two stages: (i) immediate review during image acquisition to retake photographs with blur, strong shadows, or incomplete framing; and (ii) subsequent review within the storage environment to discard images with technical defects or classification uncertainty. When diagnostic uncertainty existed, the label was revalidated with the specialist, ensuring consistency of the final dataset.

Phase 5: Study variables

The dependent variable was the leaf phytosanitary class with three categories: healthy, American leaf spot, and coffee leaf rust. The independent variable corresponded to the representation of visual features automatically obtained from the image through a convolutional feature extractor based on transfer learning, which served as input to the supervised classifiers of the hybrid model.

Phase 6: Hybrid model development (pattern recognition)

The pipeline was implemented in Python and structured as follows: (i) image input and basic normalization, (ii) feature extraction using a pretrained ResNet18 model employed as a feature extractor (transfer learning), and (iii) supervised classification. Three hybrid combinations were evaluated: ResNet18 + SVM, ResNet18 + Random Forest, and ResNet18 + XGBoost, using the same dataset and the same validation scheme to ensure comparability (Figure 2).

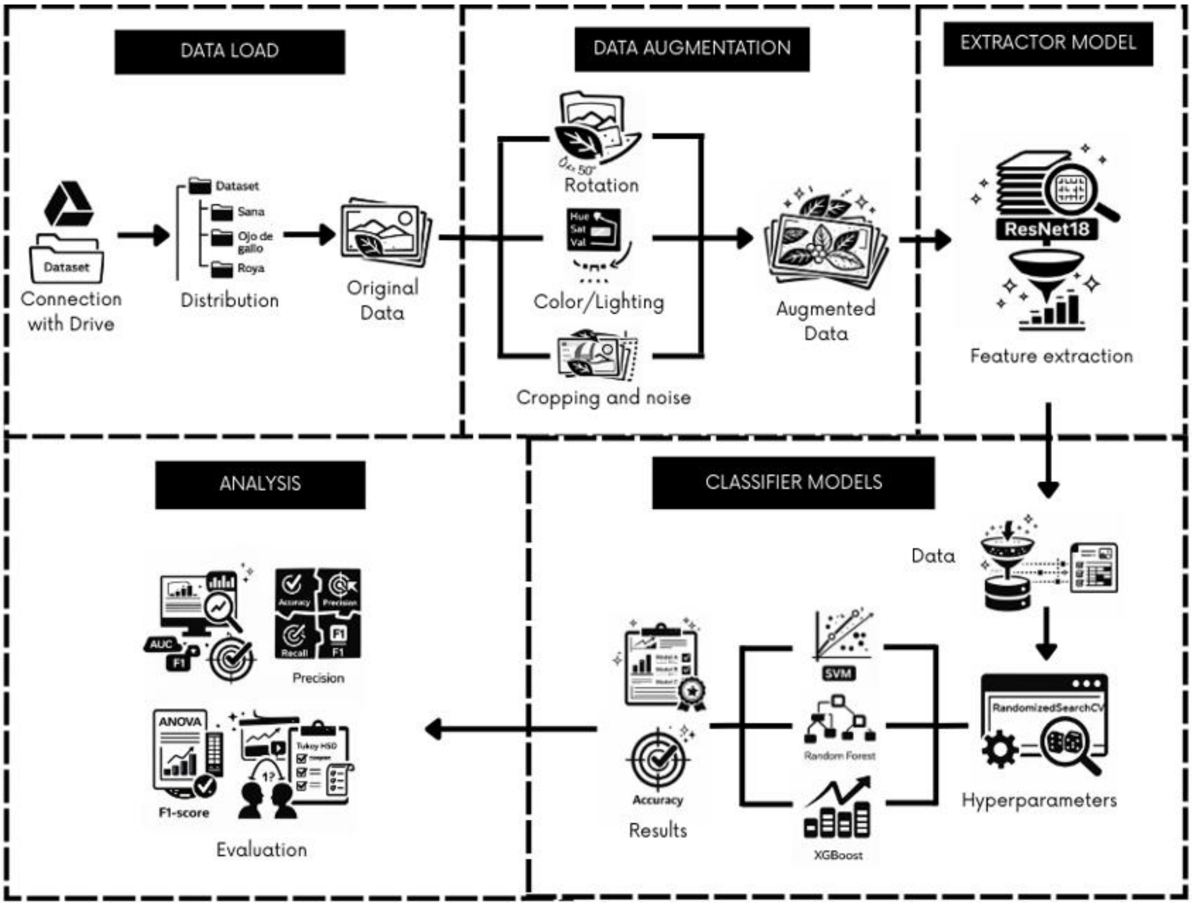

Figure 2. Flowchart of the methodological pipeline of the hybrid pattern recognition model

To ensure a fair comparison among the three hybrid models (ResNet18 + SVM, ResNet18 + Random Forest, and ResNet18 + XGBoost), the classifiers were not trained using default parameters. Hyperparameters were tuned using randomized search (RandomizedSearchCV) on the training set (80%), while the test set (20%) was kept completely independent for final evaluation. For each classifier, 30 hyperparameter configurations were explored, and stratified internal validation was applied, preserving similar proportions of the three classes. The target metric used to select the best configuration was accuracy, as it corresponds to the main metric subsequently employed in the statistical comparison (val_acc). For SVM, a pipeline incorporating standardization (StandardScaler) and (optionally) dimensionality reduction (PCA) prior to the classifier was employed. For Random Forest and XGBoost, combinations of hyperparameters related to model complexity and regularization were evaluated, maintaining the same number of iterations and the same validation scheme to ensure homogeneous conditions across models.

Phase 7: Performance evaluation and statistical analysis

Evaluation was conducted using 10-fold stratified cross-validation, preserving similar proportions of the three classes in each fold. In each fold, standard metrics were computed: accuracy, precision, recall, F1-score, and AUC, yielding 10 observations per model. To compare performance among the three hybrid models, a one-way ANOVA with fold as a blocking factor was applied; when statistical significance was detected, the Tukey HSD post-hoc test was performed to identify pairwise differences. Based on consistency and overall performance, ResNet18 + SVM was selected as the final model for the automatic classification of healthy leaves and those affected by American leaf spot and coffee leaf rust.

3. RESULTS AND DISCUSSION

The research generated, as its main contribution, a proprietary dataset of coffee leaves captured under field conditions (Saposoa, San Martín) and a comparative evaluation of three hybrid models based on feature extraction with ResNet18 and supervised classification (SVM, Random Forest, and XGBoost). The results demonstrated high performance in the identification of healthy leaves, American leaf spot, and coffee leaf rust, as well as statistically significant differences among models under cross-validation.

3.1. Dataset construction and balancing

The initial dataset consisted of 1,500 images (500 per class), as shown in Figure 3. Subsequently, through controlled offline data augmentation (three transformations per image), the dataset increased to 6,000 images, as illustrated in Figure 4, maintaining a perfectly balanced distribution of 2,000 images per class. This increase expanded visual variability under conditions similar to those encountered in the field, favoring training stability and a fair comparison among models.

A fragment of the code used for visualizing the initial leaves is shown below:

```
for cls in classes:
    path = os.path.join(segmentado_path, cls)
    imgs = [f for f in os.listdir(path)
            if f.lower().endswith(valid_ext) and "_aug" not in f.lower()][:3]
    fig, axes = plt.subplots(1, len(imgs), figsize=(12, 4))
    axes = axes if isinstance(axes, (list, np.ndarray)) else [axes]
    for ax, name in zip(axes, imgs):
        img = cv2.cvtColor(cv2.imread(os.path.join(path, name)),
cv2.COLOR_BGR2RGB)
        ax.imshow(img)
        ax.axis("off")
    plt.show()
```

Figure 3. Representative samples of the coffee leaf dataset by phytosanitary class

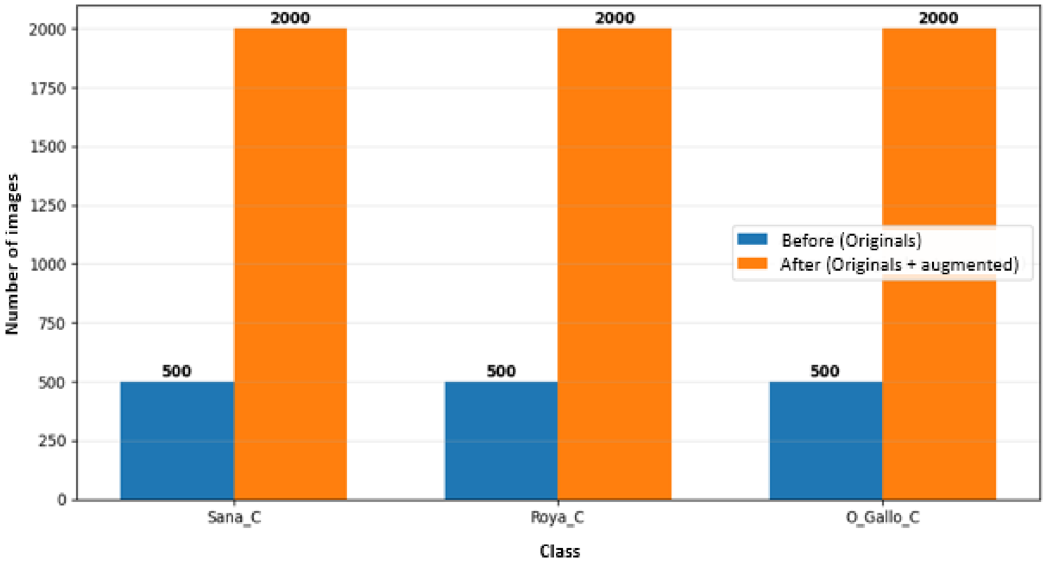

Figure 4. Distribution by class (before and after)

Before presenting the comparative results, Table 2 summarizes the final size of the dataset by class.

Table 2. Dataset distribution before and after augmentation

Class	Before augmentation	After augmentation	Percentage increase
Healthy	500	2000	150%
Leaf rust	500	2000	150%
Brown eye spot	500	2000	150%
Total	**1500**	**6000**	**150%**

It is important to clarify that data augmentation did not introduce a risk of overfitting or methodological bias in model evaluation. The *data augmentation* process was performed in an offline and controlled manner prior to dataset partitioning, and all augmented images were grouped together with their corresponding original image during the train–test split. In this way, it was ensured that no augmented version of the same leaf appeared simultaneously in both subsets, thus preventing information leakage (*data leakage*). This strategy increased the visual variability of the training set without compromising the independence of the test set, strengthening the validity of the results and the generalization capability of the evaluated models.

Regarding methodological control, dataset partitioning was carried out while avoiding augmentation-related leakage: original images and their augmented versions were grouped, ensuring that they did not appear simultaneously in the training and test sets. As a result, an 80/20 hold-out split was obtained (4,800 training; 1,200 test), and stratified cross-validation was applied on the training set.

3.2. Performance of hybrid models on the test set

All three hybrid approaches achieved high performance on unseen data (20% hold-out), with the ResNet18 + SVM model consistently outperforming the others. Table 3 presents the overall metrics obtained on the test set.

Table 3. Model performance on the TEST set (20% hold-out)

Model	Accuracy	F1-macro	AUC-macro	n_train	n_test
ResNet18 + SVM	0.974167	0.973033	0.996806	4800	1200
ResNet18 + Random Forest	0.936667	0.934056	0.987143	4800	1200
ResNet18 + XGBoost	0.958333	0.956459	0.993550	4800	1200

In comparative terms, the performance ranking on the TEST set was SVM > XGBoost > Random Forest, consistent with the trend observed during cross-validation. In particular, ResNet18 + SVM achieved the highest values of accuracy (0.9742), macro-F1 (0.9730), and macro-AUC (0.9968), suggesting a more effective combination between the embeddings extracted by ResNet18 and a classifier capable of constructing robust nonlinear decision boundaries. This behavior is expected, as SVM tends to generalize better in dense, high-dimensional feature spaces, whereas tree-based approaches (Random Forest and XGBoost) may exhibit greater sensitivity to noise and local variations when operating on embedding vectors.

Additionally, analysis of the confusion matrices allowed interpretation of the most frequent error types. Across all three models, the highest proportion of confusions occurred between coffee leaf rust and healthy leaves, indicating that certain rust cases—especially at early stages or with low severity—share visual traits with healthy leaves under the illumination and texture variability typical of field images. In the ResNet18 + SVM model, the American leaf spot class exhibited particularly stable recognition, whereas coffee leaf rust concentrated most of the errors, mainly as Rust → Healthy and, to a lesser extent, Healthy → Rust. In Random Forest and XGBoost, the same confusion pattern was observed, but with greater relative magnitude in Random Forest, which explains its lower overall performance.

From an operational perspective, the Rust → Healthy confusion constitutes the most critical error, as it implies false negatives that could delay phytosanitary interventions; in contrast, Healthy → Rust may induce false positives and unnecessary preventive decisions. Therefore, although the macro-AUC remained high in all three cases and demonstrated good multiclass discriminative capacity, the results suggest that future improvements should prioritize the separation between coffee leaf rust and healthy classes by incorporating greater diversity of mild cases, more extreme lighting conditions, and increased background variation and occlusions to reduce the errors with the greatest practical impact.

3.3. Cross-validation results and inferential comparison

In order to evaluate the stability and robustness of the models under variations in the training set, 10-fold stratified cross-validation was applied to the training data. This procedure made it possible to analyze the behavior of each model under different data partitions, simulating variability scenarios inherent to images captured under real field conditions, where factors such as illumination, leaf texture, and disease severity may vary across samples. By preserving the proportion of the three classes in each fold, cross-validation ensured a balanced evaluation and reduced dependence on a single training split, thereby strengthening the evidence of the generalization capability of the proposed approach.

To make the empirical basis used in the inferential analysis transparent, Table 4 reports the accuracy values obtained in each fold (val_acc) for the three evaluated hybrid models.

Table 4. Per-fold accuracy on the training set (10-fold cross-validation)

Fold	ResNet18+SVM	ResNet18+RF	ResNet18+XGBoost
1	0.9875	0.9563	0.9625
2	0.9771	0.9396	0.9542
3	0.9875	0.9313	0.9604
4	0.9646	0.9208	0.9396
5	0.9792	0.9396	0.9604
6	0.9750	0.9167	0.9646
7	0.9771	0.9354	0.9604
8	0.9771	0.9375	0.9563
9	0.9667	0.9208	0.9437
10	0.9563	0.9167	0.9396

Based on these values, a one-way ANOVA with fold as a blocking factor was applied, revealing statistically significant differences in performance among the models. Table 5 reports the results of the ANOVA.

Table 5. ANOVA for val_acc in CV (TRAIN) with fold as a blocking factor

Source	Sum of squares	gl	MC	F	p-value
Model	0.009396	2	0.004698	133.025489	1.647730e-11
Fold	0.002532	9	0.000281	7.965407	1.068262e-04
Residual	0.000636	18	0.000035	—	—

The analysis indicated a significant effect of the "model" factor ($p < 0.05$). The mean cross-validation accuracies (val_acc) were: SVM = 0.974792, XGBoost = 0.954167, and RF = 0.931458, maintaining the same ranking observed on the TEST set.

Given the significant result, the Tukey HSD test was applied to identify pairwise differences. Table 6 summarizes the multiple comparisons.

Table 6. Tukey HSD for val_acc (TRAIN)

Model A	Model B	Mean difference	p-value	Significant difference?
ResNet18+RF	ResNet18+SVM	0.0433	0.0000	Yes
ResNet18+RF	ResNet18+XGBoost	0.0227	0.0002	Yes
ResNet18+SVM	ResNet18+XGBoost	-0.0206	0.0006	Yes

In order to complement the inferential statistical analysis and facilitate visual interpretation of the differences among models, post hoc graphical representations based on the accuracy values obtained during cross-validation were incorporated. Figure 5 presents a boxplot of val_acc by model, which allows analysis of dispersion, stability, and the degree of overlap in performance across the 10 folds. The figure shows that the ResNet18 + SVM model concentrates the highest accuracy values and exhibits lower inter-fold variability, indicating more stable behavior under variations in the training set. In contrast, ResNet18 + Random Forest displays greater dispersion and lower central values, while ResNet18 + XGBoost shows intermediate performance.

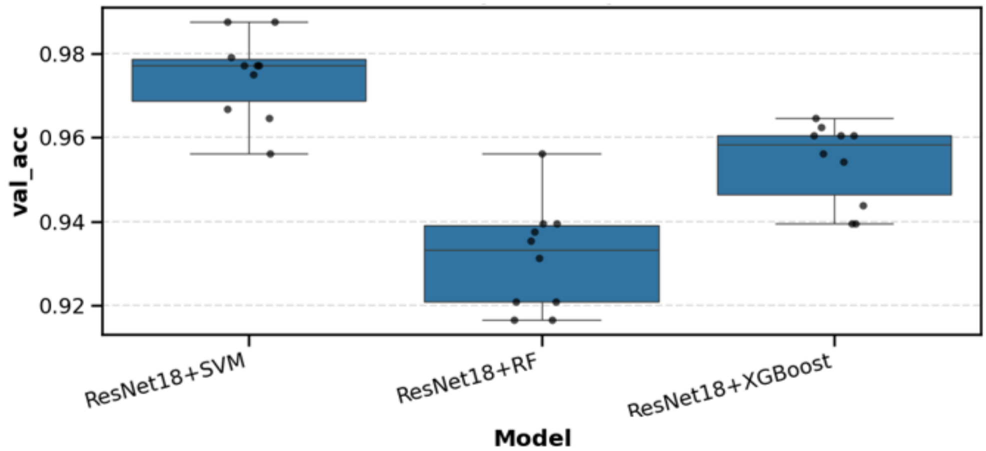

Figure 5. Distribution of val_acc in cross-validation by model

Complementarily, Figure 6 presents the mean val_acc plot under cross-validation, visually reinforcing the performance ranking identified by the ANOVA and the Tukey HSD post hoc test (ResNet18 + SVM > ResNet18 + XGBoost > ResNet18 + Random Forest). These visualizations confirm that the observed statistical differences are not only significant from an inferential standpoint but also consistent in terms of stability and generalization, which is particularly relevant given that the images were captured under real field conditions.

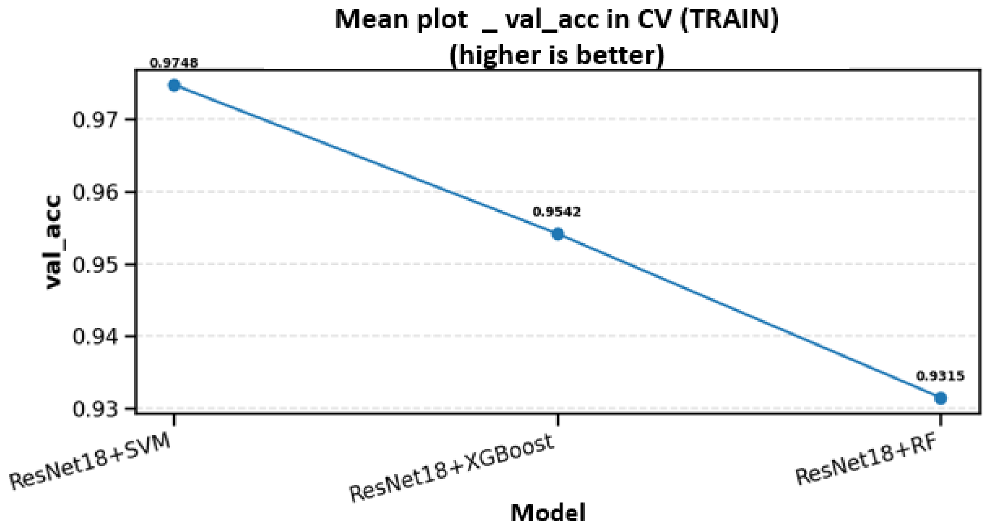

Figure 6. Mean val_acc plot under cross-validation

3.4. Detailed analysis of the selected model (ResNet18 + SVM)

Since the ResNet18 + SVM model exhibited the best overall performance and the highest consistency in both cross-validation and the test set, a detailed analysis of its classification behavior is presented below in order to deepen the interpretation of the results and assess its practical applicability in real-world scenarios.

3.4.1. Confusion matrices for training and test sets

Figure 7 shows the confusion matrix corresponding to the training set (80%). A clearly dominant diagonal can be observed, indicating an almost perfect separation among the evaluated classes (O_Gallo_C, Roya_C, and Sana_C), thus confirming the model's ability to learn discriminative visual patterns from the embeddings extracted by ResNet18.

Figure 8, in turn, presents the confusion matrix obtained on the test set (20% hold-out), composed of images never seen during training. In this scenario, the model correctly classified the vast majority of instances, with a small number of errors mainly concentrated in confusions between the Roya_C and Sana_C classes. This behavior is expected from an agronomic perspective, due to the visual similarity that healthy leaves and leaves affected by coffee leaf rust may exhibit at early stages of infection, particularly under the illumination and texture variability typical of field conditions.

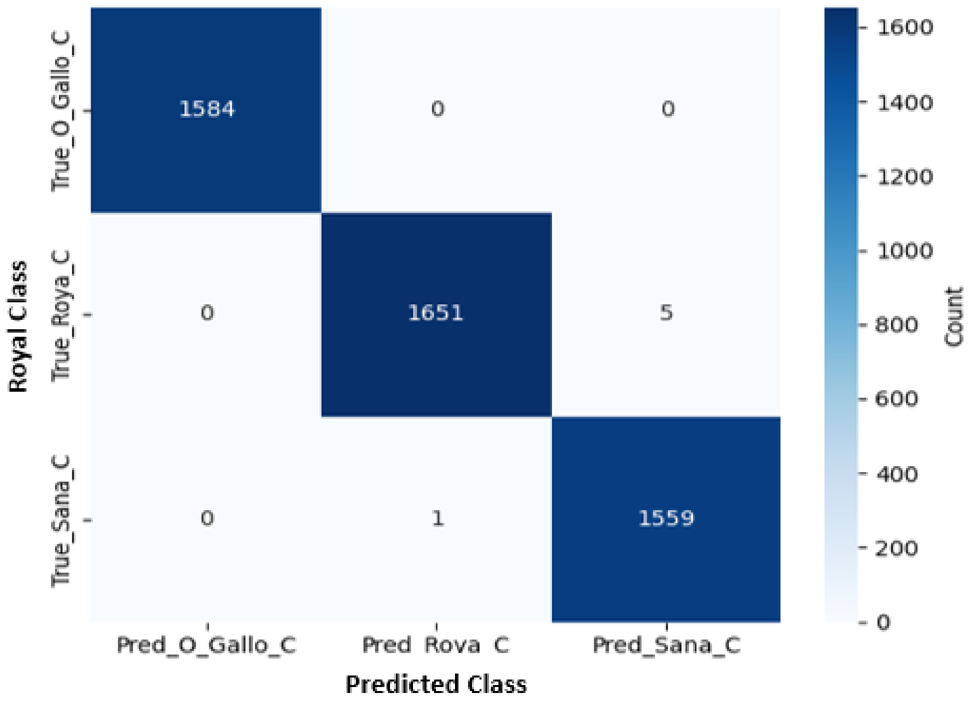

Figure 7. Confusion matrix of the ResNet18 + SVM model on the training set (80%)

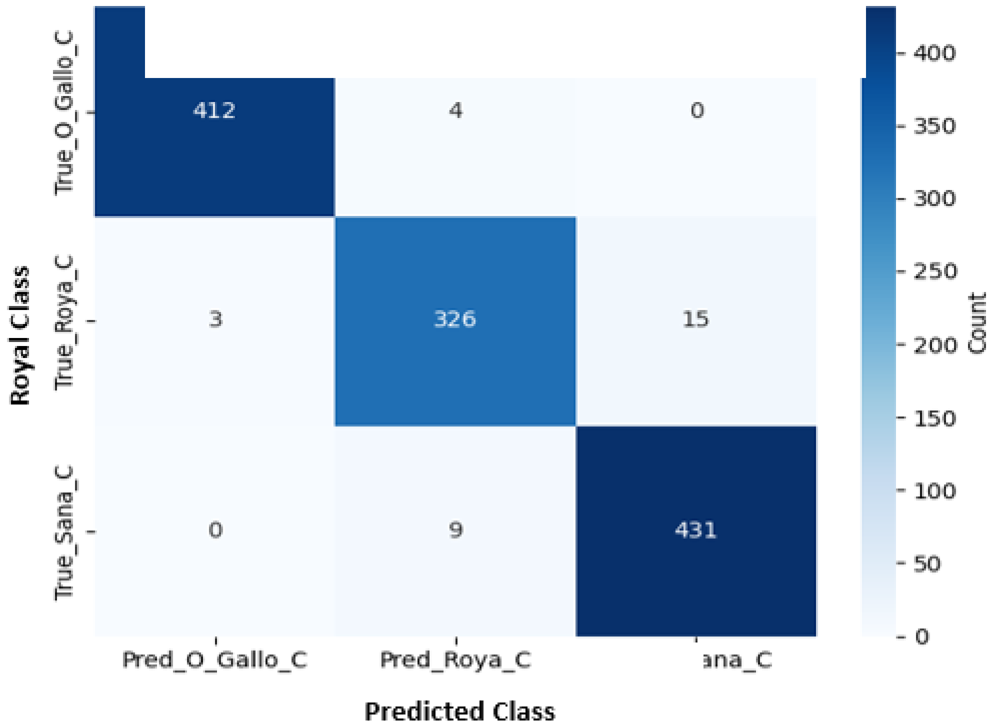

Figure 8. Confusion matrix of the ResNet18 + SVM model on the test set (20% hold-out)

3.4.2. ROC curves and discriminative capability

The discriminative capability of the model was evaluated using multiclass ROC curves under the One-vs-Rest scheme. Figure 9 shows the ROC curves corresponding to the test set, where the model achieved a macro-AUC of 0.997, with individual AUC values exceeding 0.99 for all three analyzed classes. These results confirm robust separation among categories and high system reliability in distinguishing healthy leaves, coffee leaf rust, and American leaf spot, even in the presence of variability inherent to agricultural environments.

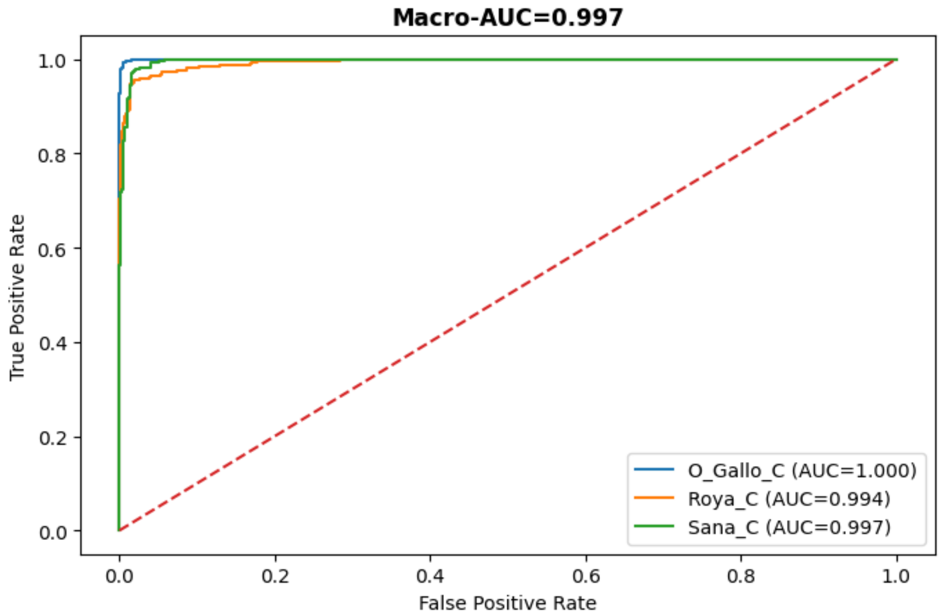

Figure 9. Multiclass ROC curves (One-vs-Rest) of the ResNet18 + SVM model on the test set (Macro-AUC = 0.997)

3.4.3. Model stability through cross-validation

The stability of the selected model was analyzed through the evolution of the accuracy and macro-F1 metrics across the 10 folds of stratified cross-validation. Figure 10 shows that both metrics remain at high values with controlled variation among folds, without abrupt performance drops. This behavior confirms that the ResNet18 + SVM model is robust to changes in the training set partition, reinforcing its generalization capability on images captured under real field conditions.

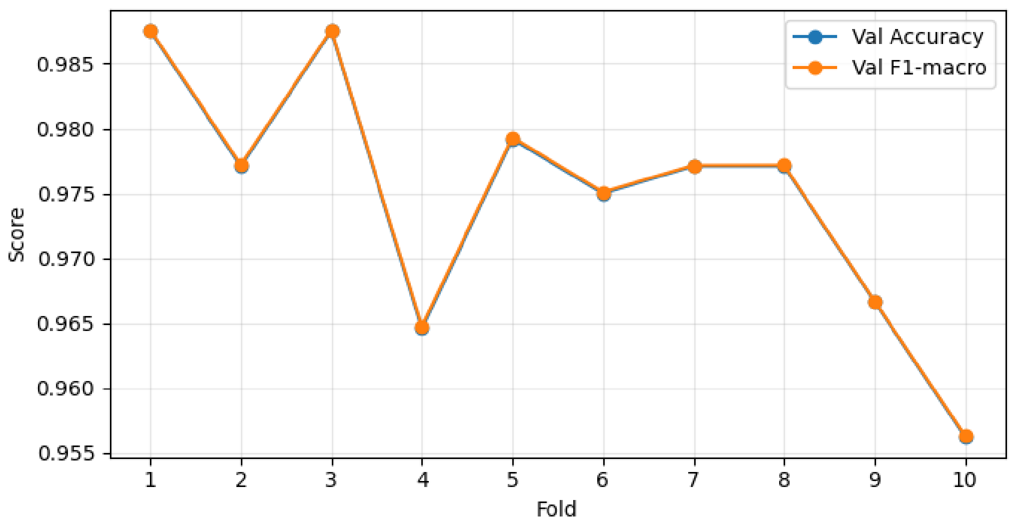

Figure 10. Evolution of accuracy and macro-F1 per fold in 10-fold cross-validation of the ResNet18 + SVM model

DISCUSSION

The obtained results demonstrated that the hybrid approach based on transfer learning and supervised classifiers enabled accurate and stable identification of healthy coffee leaves, coffee rust, and coffee leaf spot (eye spot), highlighting the ResNet18 + SVM model as the best overall-performing alternative. This finding is consistent with reports by Abuhayi & Mossa (2023)and Novtahaning et al. (2022) who showed that combining deep feature extraction with classification techniques enhances the robustness of automated diagnosis under visual variability. Likewise, recent studies have indicated that the use of deep architectures as feature extractors, combined with traditional classifiers, promotes a better balance between discrimination and model stability in coffee leaf disease classification tasks (Pham et al., 2023).

However, widely cited studies such as Atila et al. (2021), conducted on the PlantVillage dataset, reported accuracies close to 99% under controlled laboratory conditions characterized by uniform backgrounds, homogeneous illumination, and isolated leaves. While these results demonstrate the high potential of deep learning, they also reveal a recurrent limitation related to model generalization in real-world field scenarios. In contrast, the present study was based on images captured directly in coffee plantations under uncontrolled conditions, incorporating real variability in lighting, leaf texture, and disease severity. This strengthens the applied value of the proposed approach and its greater relevance for operational agricultural contexts.

Although studies such as Atila et al. (2021) and Mansouri et al. (2024) reported accuracies close to 100% using deep architectures such as EfficientNet and standardized datasets, the results of the present research were obtained from field-captured images under real lighting conditions and environmental variability. In this context, the use of ResNet18, a lighter architecture, made it possible to maintain high performance with lower computational requirements, favoring its feasibility for practical applications in agricultural environments. This advantage is consistent with studies highlighting the potential of lightweight models such as MobileNetV2 and EfficientNet-B0 for resource-constrained scenarios and deployment on low-power devices (Arif et al., 2025; Aufar & Kaloka, 2022). Nevertheless, these results should be interpreted considering certain study limitations, including the size of the dataset, its origin from a single production region, dependence on environmental conditions at the time of image capture, and the absence of external validation in other coffee-growing areas. These aspects should be addressed in future research to evaluate the spatial generalization of the model.

On the other hand, research focused on real-time detection using YOLO-based models has prioritized lesion localization with high inference speed (Adelaja & Pranggono, 2025; Fragoso et al., 2025). In contrast, the proposed approach focused on multiclasse phytosanitary classification, achieving an adequate balance between accuracy, stability, and generalization, as evidenced by cross-validation and the low performance gap between training and testing. This strategy is particularly relevant given that several mobile applications and automated diagnostic systems still exhibit limited generalization when confronted with uncontrolled image acquisition conditions (Siddiqua et al., 2022).

CONCLUSIONS

An automatic analysis model based on pattern recognition was designed and evaluated for the classification of coffee leaves into three phytosanitary conditions (healthy, rust, and eye spot). The results confirmed that the hybrid approach, integrating ResNet18 as a feature extractor with a supervised classifier, achieved high, stable, and consistent performance. Among the evaluated models, ResNet18 + SVM exhibited the best balance between accuracy, consistency, and discriminative capability, and was therefore selected as the optimal alternative.

In addition, a proprietary, structured, and balanced dataset was constructed, composed of images captured under real field conditions and validated by a specialist. Controlled data augmentation increased the dataset size without altering class balance, strengthening both training and evaluation processes. Stratified cross-validation and inferential analysis using ANOVA and the Tukey HSD post hoc test revealed statistically significant differences among the models, objectively supporting the selection of the final approach and its generalization capability within the context of the Saposoa district (San Martín).

From an applied perspective, the results demonstrate the potential of the proposed approach for integration into mobile systems to support phytosanitary diagnosis and field-based agricultural monitoring tools aimed at producers and technicians. Furthermore, the methodology could be adapted to the detection of other foliar diseases and to different crops, provided that the dataset is appropriately adjusted. In this regard, future work should focus on expanding data collection to other regions and agricultural seasons, incorporating scenarios with greater environmental variability, and evaluating deployment in mobile prototypes to assess operational performance under real usage conditions.

FINANCING

The authors did not receive sponsorship to carry out this study-article.

CONFLICT OF INTEREST

There is no conflict of interest related to the subject matter of this study.

AUTHORSHIP CONTRIBUTION

Conceptualization; Data Curation; Methodology; Project Management; Writing - original draft; Writing - revision and editing: Santa-María, J. C Formal Analysis; Investigation; Supervision: Santa-María, J. C., and Rodríguez, C. Validation was done by Rodríguez, C.

DATA AVAILABILITY

The dataset used for this study is openly available in Mendeley Data at https://doi.org/10.17632/mfpxg4y65r.1.

REFERENCES

Abade, A. S., Ferreira, P. A., & Vidal, F. de B. (2020). *Plant Diseases recognition on images using Convolutional Neural Networks: A Systematic Review*. Cornell University. https://doi.org/https://doi.org/10.48550/arXiv.2009.04365

Abdullah, H. M., Mohana, N. T., Khan, B. M., Ahmed, S. M., Hossain, M., Islam, K. S., Redoy, M. H., Ferdush, J., Bhuiyan, M. A. H. B., Hossain, M. M., & Ahamed, T. (2023). Present and future scopes and challenges of plant pest and disease (P&D) monitoring: Remote sensing, image processing, and artificial intelligence perspectives. *Remote Sensing Applications: Society and Environment*, *32*, 100996. https://doi.org/10.1016/j.rsase.2023.100996

Abuhayi, B. M., & Mossa, A. A. (2023). Coffee disease classification using Convolutional Neural Network based on feature concatenation. *Informatics in Medicine Unlocked*, *39*, 101245. https://doi.org/10.1016/j.imu.2023.101245

Adelaja, O., & Pranggono, B. (2025). Leveraging Deep Learning for Real-Time Coffee Leaf Disease Identification. *AgriEngineering*, *7*(1), 13. https://doi.org/10.3390/agriengineering7010013

Archana, R., & Jeevaraj, P. S. E. (2024). Deep learning models for digital image processing: a review. *Artificial Intelligence Review*, *57*(1), 11. https://doi.org/10.1007/s10462-023-10631-z

Arif, A., Putrawansyah, F., & Jangcik, I. (2025). Detection of Coffee Leaf Diseases Using Lightweight Deep

Learning: A Comparative Study of EfficientNet-B0 and Vision Transformer. *Ingénierie Des Systèmes d Information*, *30*(9), 2393–2404. https://doi.org/10.18280/isi.300915

Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, *61*, 101182. https://doi.org/10.1016/j.ecoinf.2020.101182

Aufar, Y., & Kaloka, T. P. (2022). Robusta coffee leaf diseases detection based on MobileNetV2 model. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*(6), 6675. https://doi.org/10.11591/ijece.v12i6.pp6675-6683

Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., Läderach, P., Anzueto, F., Hruska, A. J., & Morales, C. (2015). The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food Security*, *7*(2), 303–321. https://doi.org/10.1007/s12571-015-0446-9

Ayikpa, K. J., Ayikpa, K. J., 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

Chavarro, A. F., Renza, D., & Ballesteros, D. M. (2023). Influence of Hyperparameters in Deep Learning Models for Coffee Rust Detection. *Applied Sciences*, *13*(7), 4565. https://doi.org/10.3390/app13074565

Ehrenbergerová, L., Kučera, A., Cienciala, E., Trochta, J., & Volařík, D. (2018). Identifying key factors affecting coffee leaf rust incidence in agroforestry plantations in Peru. *Agroforestry Systems*, *92*(6), 1551–1565. https://doi.org/10.1007/s10457-017-0101-x

Fragoso, J., Silva, C., Paixão, T., Alvarez, A. B., Júnior, O. C., Florez, R., Palomino-Quispe, F., Savian, L. G., & Trazzi, P. A. (2025). Coffee-Leaf Diseases and Pests Detection Based on YOLO Models. *Applied Sciences*, *15*(9), 5040. https://doi.org/10.3390/app15095040

Julca-Otiniano, A., Alvarado-Huamán, L., Castro-Cepero, V., Borjas-Ventura, R., Gómez-Pando, L., Pereira, A. P., Nielen, S., Ingelbrecht, I., Silva, M. do C., & Várzea, V. (2024). New Races of Hemileia vastatrix Detected in Peruvian Coffee Fields. *Agronomy*, *14*(8), 1811. https://doi.org/10.3390/agronomy14081811

Mansouri, N., Guessmi, H., & Alkhalil, A. (2024). A deep learning model for detection and classification of coffee-leaf diseases using the transfer-learning technique. *International Journal of Advances in Intelligent Informatics*, *10*(3), 379. https://doi.org/10.26555/ijain.v10i3.1573

Martinez, F., Montiel, H., & Martinez, F. (2022). A Machine Learning Model for the Diagnosis of Coffee Diseases. *International Journal of Advanced Computer Science and Applications*, *13*(4). https://doi.org/10.14569/IJACSA.2022.01304110

Novtahaning, D., Shah, H. A., & Kang, J.-M. (2022). Deep Learning Ensemble-Based Automated and High-Performing Recognition of Coffee Leaf Disease. *Agriculture*, *12*(11), 1909. https://doi.org/10.3390/agriculture12111909

Pham, T. C., Nguyen, V. D., Le, C. H., Packianather, M., & Hoang, V.-D. (2023). Artificial intelligence-based solutions for coffee leaf disease classification. *IOP Conference Series: Earth and Environmental Science*, *1278*(1), 012004. https://doi.org/10.1088/1755-1315/1278/1/012004

Poma-Angamarca, R. A., Rojas, J. R., Sánchez-Rodríguez, A., & Ruiz-González, M. X. (2024). Diversity of Leaf Fungal Endophytes from Two Coffea arabica Varieties and Antagonism towards Coffee Leaf Rust. *Plants*, *13*(6), 814. https://doi.org/10.3390/plants13060814

Saavedra-Ramírez, J. (2023). Impacto social de los Proyectos de Extensión Agraria durante el 2005 al 2010 en San Martín y Amazonas, Perú. *Revista Amazónica de Ciencias Económicas*, *2*(1), e433. https://doi.org/10.51252/race.v2i1.433

Siddiqua, A., Kabir, M. A., Ferdous, T., Ali, I. B., & Weston, L. A. (2022). Evaluating Plant Disease Detection Mobile Applications: Quality and Limitations. *Agronomy*, *12*(8), 1869. https://doi.org/10.3390/agronomy12081869