Rev. Cient. Sist. Inform. 5(1), e678, doi: 10.51252/rcsi.v5i1.678

UNIVERSIDAD NACIONAL DE SAN MARTIN

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

Original article
Artículo original
Jan-Jun, 2025

Artificial vision model based on convolutional neural networks for black pod identification in cacao plantations

Modelo de visión artificial basada en redes neuronales convolucionales para identificación de mazorca negra en plantaciones de cacao

Villalobos-Culqui, Cristian[1*]
García-Rivas-Plata, Cecilia[2]
Tuesta-Hidalgo, Oscar Alejandro[3]

[1]Universidad Nacional de San Martín, Facultad de Ingeniería de Sistemas, Tarapoto, Perú
[2]Universidad Nacional Ciro Alegría, Huamachuco, La Libertad, Perú
[3]Universidad Nacional Autónoma de Alto Amazonas, Yurimaguas, Perú

Received: 9 Sept. 2024 | **Accepted:** 10 Jan. 2025 | **Published:** 20 Jan. 2025

Corresponding author*: cvcvilla013@gmail.com

How to cite this article: Villalobos-Culqui, C., García-Rivas-Plata, C. & Tuesta-Hidalgo, O.A. (2025). Artificial vision model based on convolutional neural networks for black pod identification in cacao plantations. *Revista Científica de Sistemas e Informática*, *5*(1), e678. https://doi.org/10.51252/rcsi.v5i1.678

ABSTRACT

Early detection of black pod in cocoa plantations represents a key challenge in the agricultural sector, as it affects yield and grain quality. The lack of advanced methods hinders timely identification. This study develops artificial vision models based on convolutional neural networks (CNN) to improve detection. Over nine months, we collected and labeled 1,982 images of affected pods from five plots in the Shitarillo sector, Alto Saposoa district, San Martín. We implemented YOLOv8, InceptionV3, and VGG19, applying transfer learning to optimize classification. The dataset was split into 70% for training, 20% for validation, and 10% for testing. YOLOv8 and InceptionV3 achieved an average accuracy of 79%, outperforming VGG19. Evaluation metrics, along with ANOVA and Tukey tests, confirmed that both models provided superior performance with no significant differences between them. YOLOv8 stood out for its greater robustness and accuracy, suggesting its implementation in early detection systems to optimize disease control in cocoa plantations.

Keywords: early detection, deep learning models, image analysis, automatic classification, precision agriculture

RESUMEN

La detección temprana de la mazorca negra en plantaciones de cacao representa un desafío clave en el sector agrícola, ya que afecta el rendimiento y la calidad del grano. La falta de métodos avanzados dificulta su identificación oportuna. Este estudio desarrolla modelos de visión artificial basados en redes neuronales convolucionales (CNN) para mejorar su detección. Durante nueve meses, recolectamos y etiquetamos 1982 imágenes de mazorcas afectadas en cinco parcelas del sector Shitarillo, distrito de Alto Saposoa, San Martín. Implementamos YOLOv8, InceptionV3 y VGG19, aplicando transferencia de aprendizaje para optimizar la clasificación. Dividimos los datos en 70% para entrenamiento, 20% para validación y 10% para pruebas. YOLOv8 e InceptionV3 alcanzaron una precisión promedio del 79%, superando a VGG19. Las métricas de evaluación, junto con pruebas ANOVA y Tukey, confirmaron que ambos modelos ofrecieron un desempeño superior sin diferencias significativas entre ellos. YOLOv8 destacó por su mayor robustez y exactitud, lo que sugiere su implementación en sistemas de detección temprana para optimizar el control de la enfermedad en plantaciones de cacao.

Palabras clave: detección temprana, modelos de aprendizaje profundo, análisis de imágenes, clasificación automática, agricultura de precisión

1. INTRODUCTION

Theobroma cacao L., commonly known as cocoa, is a crop of great relevance in developing countries (Palacios Bejarano et al., 2021). Its beans are used in the production of chocolates, beverages, soft drinks, ice creams, sweets, and other food products (Kumi et al., 2022; Vásquez-Cortez et al., 2024). However, one of its most common diseases, black pod rot, causes losses of up to 80% in annual production (Franco et al., 2019), making early detection and effective treatment crucial (Gyamfi et al., 2020).

In Peru, 87.3% of cocoa production comes from Ucayali (9.8%), Junín (19.9%), San Martín (39.1%), Cusco, and Amazonas (18.5%) (MIDAGRI, 2022). However, these figures show significant variations due to diseases affecting production (Cuadra et al., 2020). The San Martín region leads production with 28,984 hectares cultivated (Fachin et al., 2019), surpassing crops such as rice, coffee, and oil palm (INEI, 2022). Nevertheless, its sustainability faces major challenges, particularly due to the presence of black pod rot, caused by the fungus Phytophthora palmivora, which remains a constant concern for farmers. They must develop visual skills to identify the disease in its early stages (Fachin et al., 2019), highlighting the challenges faced by cocoa production in the region.

Our research addresses the inefficacy of traditional visual and morphological inspection methods in identifying black pod disease in cocoa. The complexity of recognizing its visible symptoms (Picon et al., 2019) and the subjectivity of the evaluation, especially when conducted by young farmers (Mashood Nasir et al., 2021; Mohammad Yazdi Pusadan et al., 2022), compromise diagnostic accuracy. Although experts can provide more reliable assessments, accessing them takes time and increases costs (Pandian et al., 2022; Singh et al., 2020), making the implementation of effective control measures more challenging (Almadhor et al., 2021; Khattak et al., 2021).

Moreover, the diversity of disease symptoms (Ahmad et al., 2023) and the influence of factors such as cocoa variety, environmental conditions, and biological interactions (Barburiceanu et al., 2021) further complicate its identification. Visual identification is challenging due to the wide variability in its morphological expression (Baculio & Barbosa, 2022; H. Li et al., 2022), which is influenced by factors such as cocoa type, environmental conditions, and biological interactions (Mzoughi & Yahiaoui, 2023). These factors, combined with the lack of precise and effective technological devices for early disease detection (Basri et al., 2020; Olofintuyi, 2022), pose a significant risk to cocoa production.

These factors lead to errors in disease identification, as phenotypic visual inspection performed by farmers or experts proves to be inaccurate (Roy et al., 2021). Consequently, cocoa quality and production are reduced (Che'Ya et al., 2022), affecting producers' competitiveness in the global market. Additionally, the indiscriminate use of fungicides (L. Li et al., 2021) can lead to plant resistance, decreasing treatment effectiveness (Jackulin & Murugavalli, 2022) and causing economic losses (Khalid et al., 2023). This issue not only impacts producers' profitability but also has implications for community health and ecological balance (Boateng et al., 2023).

The main barrier to finding a viable solution to this problem lies in the financial limitations of farmers, as the vast majority lack the means to acquire advanced technologies (Mushi et al., 2022). This situation is even more critical in developing countries, where insufficient government support for the agricultural sector and low productivity further restrict farmers' purchasing power.

Faced with this issue, we aim to determine whether there are significant differences in the accuracy of various convolutional neural network models for detecting black pod disease in cocoa plantations in the San Martín region. Our objective is to enhance early detection of this disease through the application of artificial vision models. We designed a protocol for collecting and labeling images of black pod in cocoa plantations, developed and trained convolutional neural network-based models for classification, and evaluated their performance by comparing their accuracy. Through this research, we seek to contribute to the development of technological tools that enable more efficient and precise disease identification, helping farmers improve cocoa production and sustainability.

2. MATERIALS AND METHODS

We conducted this research in the Shitarillo sector, located in the Alto Saposoa district, Huallaga province, in the San Martín department of Peru. This area belongs to the high jungle of the Peruvian Andes and is situated at an average altitude of 420 meters above sea level. The local economy primarily depends on agriculture, with cocoa being one of its main crops. The region's agroclimatic conditions, characterized by warm temperatures and high humidity, favor the spread of diseases such as black pod rot (Phytophthora palmivora), highlighting the need to develop efficient methods for early detection.

We conducted the study between October 2023 and July 2024, during which we collected and processed images of cocoa pods at different stages of infection for subsequent analysis using artificial vision models.

The study population consists of digital images of cocoa pods collected directly from the field. The images were captured using a Huawei Y9s 2019 mobile phone between 10:00 a.m. and 3:00 p.m., focusing on cocoa fruits aged 4 to 6 months from flowering.

To ensure a representative dataset, we labeled and organized the images into five categories: Stage 1, Stage 2, Stage 3, Stage 4 (corresponding to different stages of black pod infection), and healthy fruits. In total, we collected 1,982 images, distributed as follows: 400 images for each of the 4 disease stages and 382 images of healthy fruits.

Figure 1. Images Comprising the Dataset

We applied **intentional non-probabilistic sampling**, selecting images based on specific criteria of quality and representativeness of the problem under study (Olofintuyi, 2022). We split the dataset into three subsets: **70% for training, 20% for validation, and 10% for testing**, following standard practices in deep learning model evaluation (Bishop & Bishop, 2024).

Dataset Acquisition and Labeling

We captured the images using a Huawei Y9s 2019 mobile phone, set to a 12 MP (4:3) resolution, with the camera grid option activated to optimize the alignment of cocoa pods. We maintained a constant distance of 30 cm between the camera and the fruits to ensure a uniform perspective. To enhance image quality, we avoided unfavorable lighting conditions, such as direct sunlight exposure or backlighting. An agronomist engineer supervised this process to ensure the proper documentation of the samples.

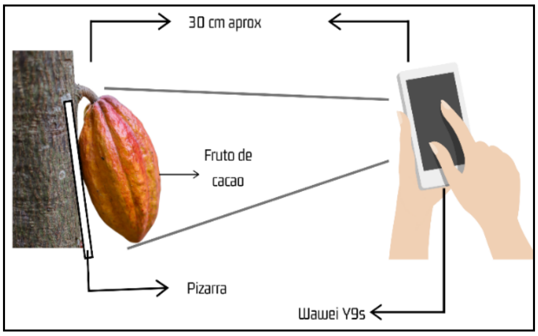

Figure 2. Schematic representation of image capture on a cocoa pod

Figure 3. Capturing images of cocoa pods infected by the disease (Stage 1)

After acquiring the images, we carried out a meticulous labeling and classification process, assigning each image to one of five categories: Stage 1, Stage 2, Stage 3, Stage 4 (corresponding to the stages of black pod disease) or Healthy Cocoa. To ensure an organized and standardized nomenclature, we implemented a Python algorithm that automated the naming assignment for each image, ensuring consistency in storage and facilitating subsequent processing.

Validation, Splitting, and Storage of the Dataset

To ensure dataset quality, we meticulously validated each image assigned to one of the five categories (Stage 1, Stage 2, Stage 3, Stage 4, and Healthy Cocoa). A plant pathology specialist agronomist reviewed and verified the correspondence of each image with its assigned category. Additionally, we conducted internal cross-checks to maintain classification consistency; any disputed labels were reevaluated before final inclusion.

We split the dataset into three subsets: training (70%), validation (20%), and testing (10%), following best practices in deep learning (Goodfellow et al., 2016). This strategy optimizes model performance by allowing it to learn from the majority of the data, adjust using an independent set, and undergo objective evaluation with entirely new images.

To preserve dataset integrity and accessibility, we implemented a secure storage protocol. Images were organized into structured folders based on their usage, with a local copy stored on a personal computer and a backup copy in Google Drive, ensuring redundancy and protection against data loss.

Subsequently, we conducted an exhaustive literature review on transfer learning, focusing on pretrained models that have proven effective in automatically classifying images into specific categories. These models leverage knowledge acquired from large datasets, enhancing disease detection in crops, minimizing the need for training from scratch, and improving generalization in new agricultural contexts.

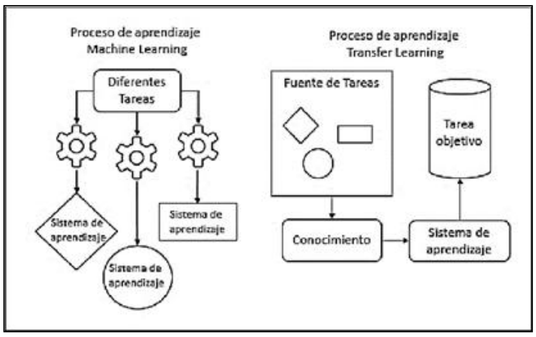

Figure 4. Contrast between traditional learning and transfer learning.

Next, we selected three convolutional neural network (CNN) models for image classification:

YOLOv8: Optimized for real-time detection, with an efficient approach to feature extraction and bounding boxes.

VGG19: Based on sequential 3x3 convolutional layers, suitable for recognizing complex hierarchical patterns.

InceptionV3: Employs Inception modules to analyze multiple feature scales simultaneously.

We implemented these models using optimization techniques such as learning rate adjustment and dropout regularization to improve generalization and prevent overfitting.

After training the models, we collected key performance metrics, including accuracy, F1-score, precision, recall, Cohen's Kappa, AUC-ROC, and confusion matrix. These metrics provided a quantitative evaluation of each model's ability to correctly classify black pod disease stages and distinguish them from healthy fruits.

Model Comparison

We compared the performance of YOLOv8, VGG19, and InceptionV3 through a classification metric-based comparative analysis. We assessed each model's ability to generalize on unseen data, analyzing their robustness in terms of accuracy, sensitivity, and computational efficiency.

To determine whether the performance differences among models were statistically significant, we applied a factorial ANOVA on the performance indicators. This test helped us identify significant variations between models and evaluate which one offered the best accuracy and stability in classification.

Visualization and analysis of results

Finally, we graphically represented the results using ROC curves, confusion matrices, and comparative diagrams, facilitating the interpretation of differences between models. This analysis allowed us to draw conclusions on which architecture is best suited for early detection of black pod disease in cocoa plantations.

3. RESULTS AND DISCUSSION

Preprocessing and labeling of black pod images in cocoa plantations

To optimize black pod detection, we applied a structured preprocessing approach to the image dataset, ensuring proper organization, cleaning, and transformation before training the models.

Dataset loading and organization

The images were divided into three subsets: 70% for training, 20% for validation, and 10% for testing. To efficiently manage data in Google Colab, we mounted the storage drive and accessed the structured folders according to their purpose.

```
from google.colab import drive
drive.mount('/content/drive')

data_dir = "/content/drive/MyDrive/Cacao_BlackPod"
train_data_dir, val_data_dir, test_data_dir = f"{data_dir}/train", f"{data_dir}/val",
f"{data_dir}/test"
```

Preprocessing and data augmentation

To enhance model generalization and mitigate overfitting, we applied normalization, scaling, and data augmentation techniques. Carefully selected transformations included rotation, translation, brightness adjustment, zoom augmentation, and horizontal flipping, following data augmentation approaches used in recent studies (Atila et al., 2021; Wang et al., 2024).

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
train_datagen = ImageDataGenerator(
    rescale=1./255,
```

```
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    brightness_range=[0.7, 1.3],
    zoom_range=0.3,
    horizontal_flip=True
)
val_test_datagen = ImageDataGenerator(rescale=1./255)
```

These configurations were experimentally adjusted to ensure optimal transformation without distorting the relevant features of black pod disease.

Dataset Loading with Image Generators.

The data were structured into batches of 32 images and loaded using image generators to facilitate model training.

```
train_dataset = train_datagen.flow_from_directory(train_data_dir, target_size=(299, 299), batch_size=32,
class_mode='categorical')
val_dataset = val_test_datagen.flow_from_directory(val_data_dir, target_size=(299, 299), batch_size=32, class_mode='categorical',
shuffle=False)
test_dataset = val_test_datagen.flow_from_directory(test_data_dir, target_size=(299, 299), batch_size=32,
class_mode='categorical', shuffle=False)
```

Preprocessing Validation

To verify the quality of the processed images, we generated a preview of the applied transformations.

```
import matplotlib.pyplot as plt
import numpy as np
batch = train_dataset.next()
images, labels = batch[0][:10], np.argmax(batch[1], axis=1)
class_labels = list(train_dataset.class_indices.keys())

plt.figure(figsize=(14, 6))
for i in range(10):
    plt.subplot(2, 5, i+1)
    plt.imshow(images[i])
    plt.xlabel(class_labels[labels[i]])
plt.show()
```

Impact of preprocessing on dataset organization

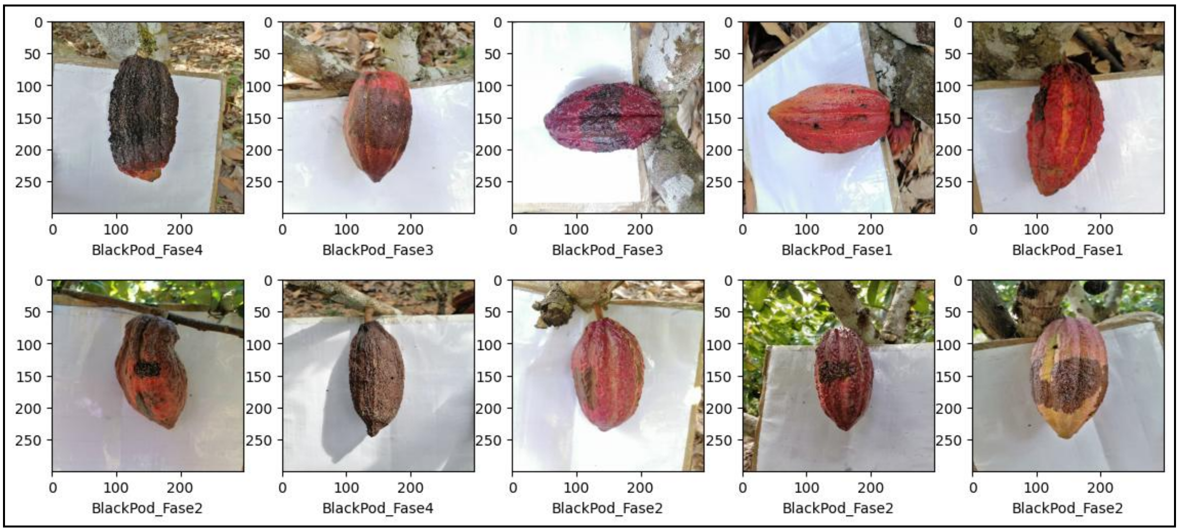

Figure 5. Preprocessed Images

The images obtained through this process were normalized, augmented, and consistently organized, ensuring compatibility with artificial vision models. This strategy, aligned with previous studies, enhances training robustness and enables more stable performance in black pod detection, similar to the approaches of (Atila et al., 2021) and (Wang et al., 2024), who implemented resizing and data augmentation techniques before processing

Training and Evaluation of Models for Black Pod Classification

To classify black pod disease, we compared three convolutional neural network (CNN) architectures: YOLOv8, InceptionV3, and VGG19.

Training with YOLOv8

We trained the YOLOv8x-cls architecture from scratch using the generated dataset.

```
from ultralytics import YOLO
model = YOLO("yolov8x-cls.pt")
results = model.train(data="/gdrive/MyDrive/Proyectos
IA/CacaoVillalobos/Datased_BlackPod",
                        epochs=20, imgsz=299)
```

Transfer Learning with InceptionV3 and VGG19

For the remaining models, we applied transfer learning, using pretrained weights on ImageNet and adapting them with additional layers.

```
from tensorflow.keras.applications import InceptionV3, VGG19
from tensorflow.keras.layers import Flatten, Dense, Dropout
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.regularizers import l2
# Cargar modelo base (solo uno se usa a la vez)
base_model = InceptionV3(weights='imagenet', include_top=False, input_shape=(299,
299, 3))
# base_model = VGG19(weights='imagenet', include_top=False, input_shape=(299, 299,
3)) # Alternativa para VGG19

# Congelar capas del modelo base
for layer in base_model.layers:
    layer.trainable = False

# Construcción del modelo
x = Flatten()(base_model.output)
x = Dense(1024, activation='relu', kernel_regularizer=l2(0.01))(x)
x = Dropout(0.2)(x)
predictions = Dense(5, activation='softmax')(x)

model = Model(inputs=base_model.input, outputs=predictions)
model.compile(optimizer=Adam(), loss='categorical_crossentropy',
metrics=['accuracy'])}}
```

Model Training

We used callbacks to save the best model and trained for 30 epochs.

```
from tensorflow.keras.callbacks import ModelCheckpoint

checkpoint = ModelCheckpoint('Cacao_InceptionV3_Best.keras', monitor='val_accuracy',
                             save_best_only=True, mode='max', verbose=1)

history = model.fit(train_dataset, epochs=30, validation_data=validation_dataset,
                    callbacks=[checkpoint])
```

Model Evaluation on the Test Set

We evaluated accuracy using the reserved test set.

```
score = model.evaluate(test_dataset)

print(f"Pérdida: {score[0]}")
```

```
print(f"Exactitud: {score[1]}")
```

Results

InceptionV3:

```
Lost:  0.9343509674072266
Accuracy:  0.7450000047683716
```

VGG19:

```
Lost:  875.97900390625
Accuracy:  0.20000000298023224
```

Performance Visualization

We displayed the evolution of accuracy during training and validation.

```
import matplotlib.pyplot as plt
plt.plot(history.history['accuracy'], label='Entrenamiento')
plt.plot(history.history['val_accuracy'], label='Validación')
plt.title('Exactitud del Modelo')
plt.xlabel('Épocas')
plt.ylabel('Exactitud')
plt.legend()
plt.show()
```

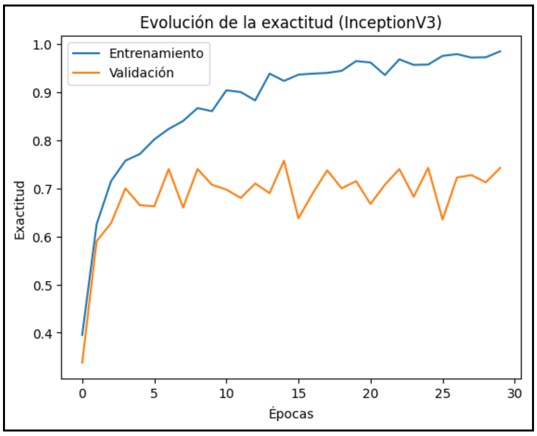

Figure 6. Training Metrics (InceptionV3)

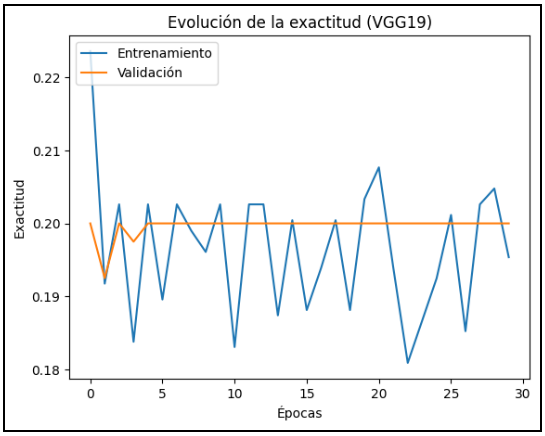

Figure 7. Training Metrics (VGG19)

The training process figures (Figure 6 and Figure 7) show that the final accuracy metric values in the training and validation sets remain stable, suggesting that the model did not experience overfitting. This indicates that the chosen architecture achieved adequate generalization to the data without excessive dependence on the training set.

In previous studies, such as Kumi et al. (2022), models for the automated detection of cocoa diseases have been developed. Their approach, based on SSD MobileNet v2, achieved a validation accuracy of 88%, outperforming the YOLOv8 model implemented in this study, which obtained 79.5%. This performance difference may be attributed to multiple factors, including the quality, quantity, and diversity of the training images used in each study. It is possible that the dataset used by Kumi et al. (2022) contained a higher volume of labeled data or higher-resolution images, which may have improved model accuracy.

However, YOLOv8 remains a viable alternative for the early detection of black pod disease, as its performance remains within a competitive margin and offers advantages in terms of computational efficiency and real-time detection. Future research could focus on optimizing the dataset, incorporating advanced data augmentation techniques, or using hybrid models to further improve classification accuracy.

Comparing the performance of the proposed artificial vision models

To evaluate and compare the performance of the trained models (YOLOv8, InceptionV3, and VGG19), we loaded each model and performed predictions on the test set.

Model loading and prediction

```
from ultralytics import YOLO
from tensorflow.keras.models import load_model
import numpy as np
# Cargar modelos
model_yolo = YOLO("./runs/classify/train10/weights/best.pt")
model_inception = load_model('Cacao_InceptionV3_Best.keras')
model_vgg19 = load_model('Cacao_VGG19_Best.keras')
# Predicciones
pred_yolo = np.argmax(model_yolo.predict(test_dataset), axis=1)
pred_inception = np.argmax(model_inception.predict(test_dataset), axis=1)
pred_vgg19 = np.argmax(model_vgg19.predict(test_dataset), axis=1)
```

Performance Evaluation

```
from sklearn.metrics import accuracy_score, f1_score, precision_score, recall_score,
cohen_kappa_score, confusion_matrix, roc_auc_score
# Obtener etiquetas reales
y_true = np.concatenate([np.argmax(batch[1], axis=1) for batch in test_dataset])
# Función para calcular métricas
def evaluar_modelo(y_true, y_pred):
    return {
        "Exactitud": accuracy_score(y_true, y_pred),
        "F1": f1_score(y_true, y_pred, average='micro'),
        "Precisión": precision_score(y_true, y_pred, average='micro'),
        "Recall": recall_score(y_true, y_pred, average='micro'),
        "Kappa de Cohen": cohen_kappa_score(y_true, y_pred),
        "ROC AUC": roc_auc_score(y_true, y_pred, multi_class="ovr")
    }
# Evaluar cada modelo
resultados = {
    "YOLOv8": evaluar_modelo(y_true, pred_yolo),
    "InceptionV3": evaluar_modelo(y_true, pred_inception),
    "VGG19": evaluar_modelo(y_true, pred_vgg19)
}
# Convertir resultados a DataFrame y mostrar
import pandas as pd
import ace_tools as tools

df_resultados = pd.DataFrame(resultados).T
tools.display_dataframe_to_user(name="Resultados de Evaluación de Modelos",
dataframe=df_resultados)
```

Confusion matrix visualization

```
import matplotlib.pyplot as plt
import seaborn as sns

cm = confusion_matrix(y_true, pred_yolo)  # Se puede cambiar a otro modelo si se requiere

plt.figure(figsize=(8,6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
plt.xlabel('Predicción')
plt.ylabel('Referencia')
plt.title('Matriz de Confusión - YOLOv8')
plt.show()
```

Which results in the following outcome:

Table 1. Comparison of performance metrics between YOLOv8, InceptionV3, and VGG19 in black pod classification.

INDICATOR	Yolov8	Inceptionv3	VGG19
Accuracy	0.7902	0.7321	0.1920
F1-Score	0.7902	0.7321	0.1920
Precision	0.7902	0.7321	0.1920
Recall	0.7902	0.7321	0.1920
Cohen's Kappa	0.7315	0.6535	0.0000
ROC AUC	0.854	0.835	0.50

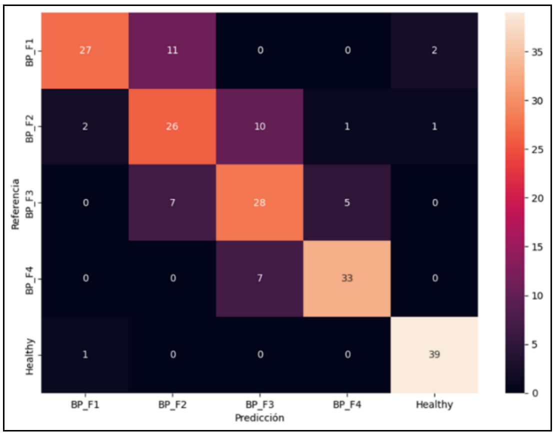

Figure 8. Confusion Matrix (YOLOv8)

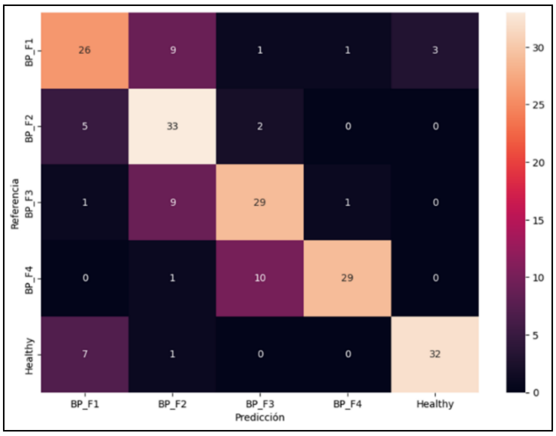

Figure 9. Confusion Matrix (InceptionV3)

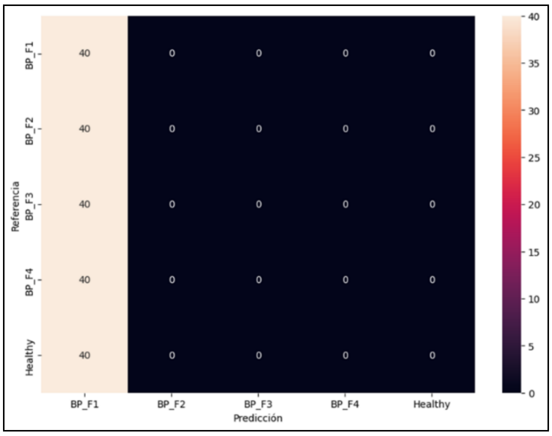

Figure 10. Confusion Matrix (VGG19)

The confusion matrix results show that the YOLOv8 model achieved the highest accuracy (79.02%), efficiently classifying all stages of black pod disease and demonstrating strong recognition of healthy fruits (97.5% accuracy in the Healthy category). A slight confusion was observed between BP_F1 and BP_F2 (11 errors), suggesting that the characteristics of these early stages may appear similar to the model. InceptionV3, with an accuracy of 73.21%, also demonstrated acceptable classification performance, though it exhibited a higher number of errors in identifying healthy fruits (7 misclassifications in the Healthy category) and struggled to distinguish BP_F3 from BP_F4. In contrast, VGG19 proved ineffective (19.20% accuracy), as it collapsed all predictions into the BP_F1 category, indicating that it failed to generalize the information across different classes correctly.

These findings are consistent with previous research on crop disease detection using convolutional neural networks (Kumi et al., 2022; Mohammad Yazdi Pusadan et al., 2022; Olofintuyi, 2022). In a study by Baculio & Barbosa (2022) on the classification of cocoa brown rot, YOLOv5 outperformed EfficientNet with 81% accuracy, demonstrating that YOLO architectures are more efficient for real-time agricultural classification problems. However, model accuracy heavily depends on dataset quality, which may explain the difference between this study and our findings. While YOLOv8 proved to be the most efficient model, future improvements could focus on increasing the amount of training data and applying advanced data augmentation techniques to enhance accuracy in the early stages of the disease, where the highest confusion was observed.

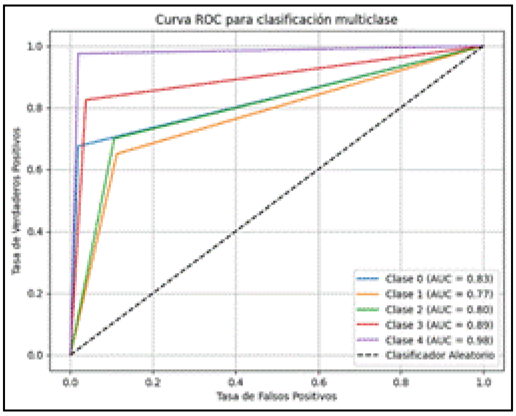

Figure 11. AUC ROC (YOLOv8)

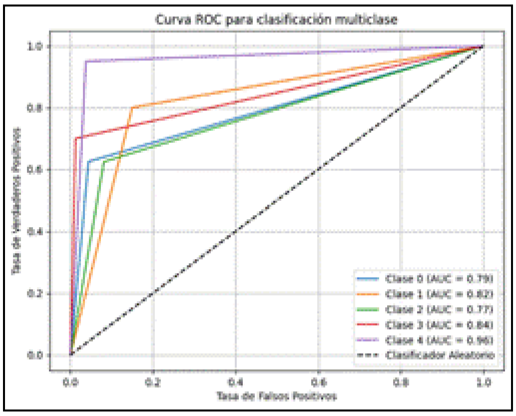

Figure 12. AUC ROC (InceptionV3)

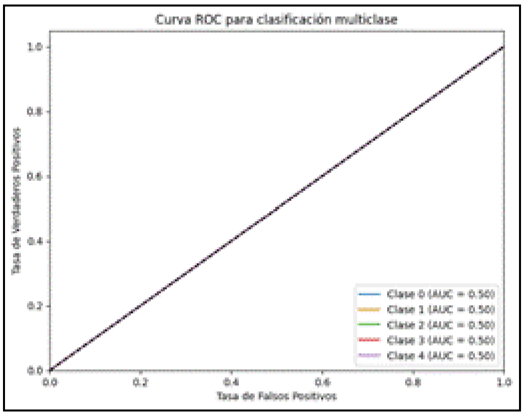

Figure 13. AUC ROC (VGG19)

The ROC curves show that YOLOv8 achieved the best discriminative ability among classes, with AUC values ranging from 0.77 to 0.98, indicating high accuracy in classifying black pod disease and healthy fruits. InceptionV3 demonstrated a similar performance, with AUC values between 0.77 and 0.96, though with a slight reduction in discriminative ability for some classes compared to YOLOv8. In contrast, VGG19 recorded an AUC of 0.50 across all classes, indicating performance equivalent to a random classifier, confirming its inability to differentiate between categories.

The AUC values obtained for YOLOv8 and InceptionV3 align with previous studies, such as Basri et al., (2020), where CNN-based models achieved AUC values above 0.85 in crop disease classification for coffee plantations. The difference between YOLOv8 and InceptionV3 may be attributed to YOLOv8's ability to better capture spatial features in real-time, making it more efficient for agricultural scenarios. On the other hand, VGG19's poor performance suggests that its architecture is not suitable for this task, as it failed to learn relevant patterns from the images. It is recommended to optimize the models by applying data augmentation and hyperparameter tuning to improve classification performance in classes with lower AUC values.

Statistical Analysis of Model Performance

To evaluate the performance variability of each model, the test set was divided into seven subsets, and each model was executed on each subset. Table 2 presents the results obtained for each metric:

Table 2. Descriptive Statistics of Performance Metric

Model	N	Media	D. Est.	Er. Est.	IC 95% (L.I)	IC 95% (L.S)	Min	Max
YOLOv8	7	**0.7902**	0.09149	0.03458	0.7056	0.8748	0.69	0.91
InceptionV3	7	**0.7321**	0.05945	0.02247	0.6772	0.7871	0.63	0.81
VGG19	7	**0.1920**	0.04917	0.01859	0.1465	0.2374	0.13	0.25
Total	21	**0.5714**	0.28369	0.06191	0.4423	0.7006	0.13	0.91

The average values indicate that YOLOv8 achieved the highest accuracy (79.02%), followed by InceptionV3 (73.21%), while VGG19 exhibited significantly lower performance (19.20%). Additionally, the 95% confidence intervals reveal that YOLOv8 has a higher variability margin compared to InceptionV3, yet it consistently remains within a higher performance range.

ANOVA Test and Tukey Post-hoc Analysis

To determine whether the differences between models were statistically significant, a one-way ANOVA was applied to the results of accuracy, precision, recall, F1-score, and Cohen's Kappa:

Table 3. ANOVA Test Results

Variable	Sum of squares	gl	Mean square	F	Sig. (p-valor)
Accuracy	1.524	2	0.762	159.575	**0.000**
F1-score	1.524	2	0.762	159.575	**0.000**
Precision	1.524	2	0.762	159.575	**0.000**
Recall	1.524	2	0.762	159.575	**0.000**
Cohen's Kappa	2.259	2	1.130	161.610	**0.000**

Since in all evaluated metrics p-value < 0.01, we conclude that there are statistically significant differences among at least one of the evaluated models.

To identify which models, exhibit significant differences, a Tukey post-hoc test was applied, and the results are presented in Table 4:

Table 4. Tukey Test Results (Accuracy)

Comparison	Mean Difference	Std. Error	Sig. (p-valor)	IC 95% (Lim Inf)	IC 95% (Lim Sup)
YOLOv8 - InceptionV3	0.05804	0.03693	**0.283**	-0.0362	0.1523
YOLOv8 - VGG19	0.59821**	0.03693	**0.000**	0.5040	0.6925
InceptionV3 - VGG19	0.54018**	0.03693	**0.000**	0.4459	0.6344

The results indicate that YOLOv8 and InceptionV3 do not exhibit significant differences (p = 0.283), suggesting that both models have similar performance. However, VGG19 is significantly inferior to both models (p < 0.001), confirming its ineffectiveness in black pod classification.

The statistical analysis demonstrates that YOLOv8 and InceptionV3 are viable models for black pod detection, though YOLOv8 shows superior accuracy and consistency. The absence of significant differences between these two models suggests that both can be used in real-world applications, although YOLOv8 offers better generalization due to its optimized structure for real-time detection.

Comparing these results with previous research, (Kumi et al., 2022) reported that SSD MobileNet achieved 80% accuracy in cocoa disease classification, a value very close to the 79% accuracy of YOLOv8 in this study. The slight difference could be explained by variations in dataset quality and size used during training. Additionally, the high dispersion in Cohen's Kappa values for InceptionV3 suggests that this model may be more sensitive to data variations, which could affect its implementation in environments with greater environmental variability.

CONCLUSIONS

We successfully improved the early detection of black pod disease in cocoa plantations in the San Martín region through the artificial vision models developed in this research, two of which demonstrated strong performance in the early identification of this disease.

We successfully and thoroughly collected and labeled a dataset of 1,982 images of black pod disease in cocoa plantations from five plots located in the Shitarillo sector, Alto Saposoa district, Huallaga province, San Martín region.

We built three artificial vision models using advanced techniques, including convolutional neural networks for the YOLOv8-based model and transfer learning for the InceptionV3 and VGG19-based models. Two of these models proved effective in classifying black pod images segmented into five categories, which were further divided into training, validation, and test groups to ensure a structured training process.

After building and training the three models, we evaluated and compared their performance using classification performance indicators applied to seven data subsets. These values were subjected to an ANOVA test, which determined that at least one model had significant differences compared to the others. The Tukey test confirmed that the VGG19-based model had a significantly lower performance than the other two models. However, the remaining two models showed no significant differences between them.

FINANCING

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

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest related to the development of the study.

AUTHORSHIP CONTRIBUTION

Conceptualization; Data curation; Formal analysis; Research; Methodology; Visualization; Writing - original draft; Writing - review and editing: Villalobos-Culqui, Cristian, García-Rivas-Plata, Cecilia, Tuesta-Hidalgo, Oscar Alejandro.

REFERENCES

Ahmad, A., Saraswat, D., & El Gamal, A. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology*, *3*, 100083. https://doi.org/10.1016/j.atech.2022.100083

Almadhor, A., Rauf, H., Lali, M., Damaševičius, R., Alouffi, B., & Alharbi, A. (2021). AI-Driven Framework for Recognition of Guava Plant Diseases through Machine Learning from DSLR Camera Sensor Based High Resolution Imagery. *Sensors*, *21*(11), 3830. https://doi.org/10.3390/s21113830

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

Baculio, N. G., & Barbosa, J. B. (2022). An Objective Classification Approach of Cacao Pods using Local Binary Pattern Features and Artificial Neural Network Architecture (ANN). *Indian Journal of Science and Technology*, *15*(11), 495–504. https://doi.org/10.17485/IJST/v15i11.60

Barburiceanu, S., Meza, S., Orza, B., Malutan, R., & Terebes, R. (2021). Convolutional Neural Networks for Texture Feature Extraction. Applications to Leaf Disease Classification in Precision Agriculture. *IEEE Access*, *9*, 160085–160103. https://doi.org/10.1109/ACCESS.2021.3131002

Basri, Tamin, R., Karim, H. A., Indrabayu, & Areni, I. S. (2020). Mobile image processing application for cacao's fruits pest and disease attack using deep learning algorithm. *ICIC Express Letters*, *14*. https://doi.org/https://doi.org/10.24507/icicel.14.10.1025

Bishop, C. M., & Bishop, H. (2024). *Deep Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-031-45468-4

Boateng, K. O., Dankyi, E., Amponsah, I. K., Awudzi, G. K., Amponsah, E., & Darko, G. (2023). Knowledge, perception, and pesticide application practices among smallholder cocoa farmers in four Ghanaian cocoa-growing regions. *Toxicology Reports*, *10*, 46–55. https://doi.org/10.1016/j.toxrep.2022.12.008

Che'Ya, N. N., Mohidem, N. A., Roslin, N. A., Saberioon, M., Tarmidi, M. Z., Arif Shah, J., Fazlil Ilahi, W. F., & Man, N. (2022). Mobile Computing for Pest and Disease Management Using Spectral Signature Analysis: A Review. *Agronomy*, *12*(4), 967. https://doi.org/10.3390/agronomy12040967

Cuadra, Y. M. L., Rodríguez, M. Y. C., & Vega, Y. L. C. (2020). *El cacao peruano y su impacto en la economía nacional*.

Fachin, G., Pinedo, K., Vásquez, J., Flores, E., Doria, M., Alvarado, J., Koch, C., & Bellido, J. J. (2019). Factores ambientales y su relación con la incidencia de carmenta foraseminis (Busck) eichlin (Lepidóptera : Sesiidae) en frutos de theobroma cacao "cacao" en San Martín, Perú. *Boletín Científico Centro de Museos Museo de Historia Natural*, *23*(2), 133–145. https://doi.org/10.17151/bccm.2019.23.2.6

Franco, A. J. D., Merca, F. E., Rodriguez, M. S., Balidion, J. F., Migo, V. P., Amalin, D. M., Alocilja, E. C., & Fernando, L. M. (2019). DNA-based electrochemical nanobiosensor for the detection of Phytophthora palmivora (Butler) Butler, causing black pod rot in cacao (Theobroma cacao L.) pods. *Physiological and Molecular Plant Pathology*, *107*, 14–20. https://doi.org/10.1016/j.pmpp.2019.04.004

Gyamfi, A., Iddrisu, S. A.-J., & Adegbola, O. (2020). Machine Learning-Based Cocoa E-Health System. *2020 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges(51275)*, 1–7. https://doi.org/10.1109/CMI51275.2020.9322689

INEI. (2022). *Informe Técnico No 01 - Marzo 2022 Indicador de la Actividad Productiva Departamental*.

Jackulin, C., & Murugavalli, S. (2022). A comprehensive review on detection of plant disease using machine learning and deep learning approaches. *Measurement: Sensors*, *24*, 100441. https://doi.org/10.1016/j.measen.2022.100441

Khalid, M., Sarfraz, M., Iqbal, U., Aftab, M., Niedbała, G., & Rauf, H. (2023). Real-Time Plant Health Detection Using Deep Convolutional Neural Networks. *Agriculture*, *13*(2), 510. https://doi.org/10.3390/agriculture13020510

Khattak, A., Asghar, M. U., Batool, U., Asghar, M. Z., Ullah, H., Al-Rakhami, M., & Gumaei, A. (2021). Automatic Detection of Citrus Fruit and Leaves Diseases Using Deep Neural Network Model. *IEEE Access*, *9*, 112942–112954. https://doi.org/10.1109/ACCESS.2021.3096895

Kumi, S., Kelly, D., Woodstuff, J., Lomotey, R. K., Orji, R., & Deters, R. (2022). Cocoa Companion: Deep Learning-Based Smartphone Application for Cocoa Disease Detection. *Procedia Computer Science*, *203*, 87–94. https://doi.org/10.1016/j.procs.2022.07.013

Li, H., Shi, H., Du, A., Mao, Y., Fan, K., Wang, Y., Shen, Y., Wang, S., Xu, X., Tian, L., Wang, H., & Ding, Z. (2022). Symptom recognition of disease and insect damage based on Mask R-CNN, wavelet transform, and F-RNet. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.922797

Li, L., Zhang, S., & Wang, B. (2021). Plant Disease Detection and Classification by Deep Learning—A Review. *IEEE Access*, *9*, 56683–56698. https://doi.org/10.1109/ACCESS.2021.3069646

Mashood Nasir, I., Bibi, A., Hussain Shah, J., Attique Khan, M., Sharif, M., Iqbal, K., Nam, Y., & Kadry, S. (2021). Deep Learning-based Classification of Fruit Diseases: An Application for Precision Agriculture. *Computers, Materials & Continua*, *66*(2), 1949–1962. https://doi.org/10.32604/cmc.2020.012945

MIDAGRI. (2022). *Commodities Cacao_ abr-jun 2022*. https://repositorio.midagri.gob.pe/bitstream/20.500.13036/1363/1/Commodities%20Cacao_%20abr-jun%202022.pdf

Mohammad Yazdi Pusadan, Syahrullah, Merry, & Ahmad Imam Abdullah. (2022). k-Nearest Neighbor and Feature Extraction on Detection of Pest and Diseases of Cocoa. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(3), 471–480. https://doi.org/10.29207/resti.v6i3.4064

Mushi, G. E., Di Marzo Serugendo, G., & Burgi, P.-Y. (2022). Digital Technology and Services for Sustainable Agriculture in Tanzania: A Literature Review. *Sustainability*, *14*(4), 2415. https://doi.org/10.3390/su14042415

Mzoughi, O., & Yahiaoui, I. (2023). Deep learning-based segmentation for disease identification. *Ecological Informatics*, *75*, 102000. https://doi.org/10.1016/j.ecoinf.2023.102000

Olofintuyi, S. S. (2022). Early Cocoa Blackpod Pathogen Prediction with Machine Learning Ensemble Algorithm based on Climatic Parameters. *Journal of Information and Organizational Sciences*, *46*(1), 1–14. https://doi.org/10.31341/jios.46.1.1

Palacios Bejarano, C. A., Afanador Kafuri, L., & Pardo García, J. M. (2021). Identificación de aislados de Phytophthora spp. obtenidos de cultivos de cacao en Antioquia, Colombia. *Acta Agronómica*, *70*(1), 17–26. https://doi.org/10.15446/acag.v70n1.70619

Pandian, J., Kumar, V., Geman, O., Hnatiuc, M., Arif, M., & Kanchanadevi, K. (2022). Plant Disease Detection Using Deep Convolutional Neural Network. *Applied Sciences*, *12*(14), 6982. https://doi.org/10.3390/app12146982

Picon, A., Seitz, M., Alvarez-Gila, A., Mohnke, P., Ortiz-Barredo, A., & Echazarra, J. (2019). Crop conditional Convolutional Neural Networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. *Computers and Electronics in Agriculture*, *167*, 105093. https://doi.org/10.1016/j.compag.2019.105093

Roy, S., Ray, R., Dash, S. R., & Giri, M. K. (2021). Plant Disease Detection Using Machine Learning Tools With an Overview on Dimensionality Reduction. In *Data Analytics in Bioinformatics* (pp. 109–144). Wiley. https://doi.org/10.1002/9781119785620.ch5

Singh, V., Sharma, N., & Singh, S. (2020). A review of imaging techniques for plant disease detection. *Artificial Intelligence in Agriculture*, *4*, 229–242. https://doi.org/10.1016/j.aiia.2020.10.002

Vásquez-Cortez, L. H., Rojas-Sánchez, J., Pazmiño-Pérez, Á., & Rodríguez-Cevallos, S. L. (2024). Desarrollo de una bebida energizante a base de cacao y guayusa. *Revista Peruana de Investigación Agropecuaria*, *3*(2), e66. https://doi.org/10.56926/repia.v3i2.66

Wang, Y., Rajkumar Dhamodharan, U. S., Sarwar, N., Almalki, F. A., Naith, Q. H., R, S., & D, M. (2024). A hybrid approach for rice crop disease detection in agricultural IoT system. *Discover Sustainability*, *5*(1), 99. https://doi.org/10.1007/s43621-024-00285-4