YOLO and RF-DETR architectures

for early disease diagnosis in Solanum tuberosum L

Manolo Sebastián Muñoz-Espinoza[1]

[1]Universidad Técnica de Ambato, Facultad de Ciencias Agropecuarias, Ambato, Ecuador,
https://orcid.org/0000-0001-5897-1211
mmunoz@uta.edu.ec

DOI: 10.61854/rccedia.v1n1.005
Sent: February 12, 2026
Approved: April 6, 2026
Published: Tuesday, July 14, 2026

Abstract

Three object detectors for post-harvest diagnosis in tubers were compared under a unified protocol: two YOLO-family architectures and a transformer-based detector (RF-DETR). We used a six-class public dataset (five diseases plus healthy). Lesions were manually annotated with bounding boxes in Roboflow and augmented with geometric/photometric transforms. Annotations were converted to COCO for RF-DETR. We enforced identical splits (train/val/test = 792/56/57), the same input size (640 px), and an identical training recipe (60 epochs with early stopping). Evaluation on the held-out test set reported precision, recall, mAP@0.50, and mAP@0.50:0.95. On test images, YOLOv12 achieved P = 78.72%, R = 96.86%, mAP@0.50 = 79.36%, mAP@0.50:0.95 = 73.26%; YOLOv26 reached P = 79.34%, R = 96.64%, mAP@0.50 = 78.48%, mAP@0.50:0.95 = 73.55% (experimental logs). RF-DETR delivered P = 70.86%, R = 88.39%, mAP@0.50 = 85.61%, and mAP@0.50:0.95 = 82.64%. For high-sensitivity screening (high recall), the YOLO architecture is more convenient, while RF-DETR offers strict localization (superior mAP at 0.50–0.95) and simplifies post-processing by eliminating the need for an NMS. A two-stage pipeline (YOLO → RF-DETR) balances speed with spatial accuracy. Future work includes per-class score calibration and multimodal fusion with NIR/HSI for subepidermal symptoms.

Keywords:
precision agriculture, deep learning, data augmentation, object localization, computer vision.

Introduction

Potato (Solanum tuberosum L.) is one of the world's most important food crops and ranks third in global consumption (Nagel et al., 2022). In 2022, global production exceeded 375 million tons, with China and India as the leading producers (95.5 and 56 million tons, respectively) (Guillemette et al., 2024; Solankey, 2025). Its importance is supported not only by its nutritional value and culinary versatility but also by its strategic role in food security and the rural economy, particularly in developing countries (Moreno Mayhuire & Herrera Quispe, 2023). However, productivity and postharvest quality are affected by fungal, bacterial, and physiological diseases that directly impact tubers (Keithellakpam et al., 2025).

Among the economically significant diseases are blackleg (Pectobacterium spp.), pink rot (Phytophthora erythroseptica), dry rot (Fusarium spp.), common scab (Streptomyces scabies), and black scurf (Rhizoctonia solani) (Sharma et al., 2024; Upeniece et al., 2025). These diseases can cause postharvest losses ranging from 30% to 50% when they are not detected and managed in a timely manner (Aleinzi, 2025).

Early diagnosis is an essential component of the agricultural value chain. However, methods based on expert visual inspection are often subjective, slow, costly, and dependent on human expertise, limiting their scalability and reproducibility (Qin et al., 2024). In contrast, artificial intelligence (AI) approaches, particularly deep learning applied to computer vision, have emerged as valuable tools in precision agriculture (Jayanthi & Kumar, 2024). Object detection models offer advantages over classification approaches by enabling not only the identification of disease presence but also its localization on the surface of the tuber (Ma et al., 2022).

In this context, Ultralytics YOLO (You Only Look Once) has demonstrated a balance between accuracy, speed, and computational efficiency, motivating its adoption in agricultural applications (Divyanth et al., 2024; Geoffrey Manurung et al., 2024; LIU, 2024; Palacio Betancur & Bolaños Martínez, 2025; Redmon, 2016). Previous versions, such as YOLOv5, YOLOv8, and YOLOv12, have shown promising results in the detection of pests and diseases affecting crops such as tomato, strawberry, and potato (He et al., 2024; Manolo Muñoz-Espinoza et al., 2026; J. Zhang et al., 2024).

YOLO26 is a recent real-time object detection model designed for edge devices and low energy consumption. It employs an optimized architecture that reduces computational complexity, facilitating agile, lightweight, and accessible deployment (Jocher & Qiu, 2026; Sapkota et al., 2025; Yurdakul et al., 2026).

Similarly, RF-DETR is a real-time transformer-based object detection architecture belonging to the DETR (Detection Transformer) family. It is adaptable to different domains and datasets of varying sizes and is intended for applications requiring high speed and accuracy, even in environments with limited computational resources (e.g., edge devices or low-latency applications) (Robinson et al., 2026).

Despite these advances, several challenges remain: (1) the scarcity of local, public, and annotated datasets representing real field or postharvest conditions; (2) the need to adapt models to contexts with limited samples through robust data augmentation; and (3) rigorous validation in multiclass detection tasks involving visually similar classes.

Revista Científica CEDIA - ISSN: 3151-8435

To compare complementary design priorities, this study analyzed two representative detectors from the YOLO family—which prioritize speed and ease of deployment—and expanded the scope by incorporating the transformer-based detector RF-DETR, which formulates detection as set prediction based on queries and one-to-one assignment, reducing dependence on anchors and intensive post-processing (Robinson et al., 2026). To ensure experimental fairness, the three detectors were trained using the same images, identical training/validation/test splits, and the same input resolution, while maintaining a unified evaluation protocol. The annotations were converted to the COCO format for training the DETR-based model.

The proposed approach employs an open potato tuber dataset (Faria et al., 2023), containing six classes (five diseases and one healthy condition), annotated in Roboflow (Roboflow: Computer Vision Tools for Developers and Enterprises, 2025) and expanded through a data augmentation protocol designed to simulate realistic variability in orientation, illumination, and noise.

The primary objective was to comparatively evaluate the performance of YOLO-family detectors and transformer-based detectors (RF-DETR) for the early detection of diseases in potato tubers, with the aim of developing an automated, accurate, and scalable tool that can be integrated into real-time monitoring systems and deployed in resource-constrained environments.

Contributions

- A comparative benchmark between CNN-based detectors from the YOLO family and a transformer-based detector (RF-DETR), using real agricultural potato images comprising six classes characterized by high visual variability and uncontrolled acquisition conditions.
- A reproducible annotation and COCO conversion pipeline, with clearly defined comparability criteria, including identical data splits, uniform input resolution, and consistent data augmentation strategies.
- A homogeneous comparative evaluation focused on analyzing model behavior, complementing standard metrics (mAP@0.5 and mAP@0.5:0.95) with precision, recall, and confusion matrices, enabling the identification of systematic errors and domain-specific challenges in agricultural applications.

 Revista Científica CEDIA - ISSN: 3151-8435

Methods

The study presents a comparative experimental design involving three object detectors (YOLOv12, YOLOv26, and RF-DETR) evaluated under a common protocol. The complete processing workflow consisted of: (1) acquisition and manual annotation of lesions in 377 original images; (2) stratified splitting into training (70%), validation (15%), and test (15%) sets; (3) augmentation applied exclusively to the training set; (4) conversion of annotations to COCO format for RF-DETR; (5) independent training of each model using the same input resolution (640 px), identical data splits, and identical evaluation criteria; and (6) final evaluation on the test set using standardized metrics (precision, recall, mAP@0.50, and mAP@0.50:0.95).

Original dataset and inclusion /exclusion criteria

The original dataset was obtained from the Kaggle platform (Faria et al., 2023).

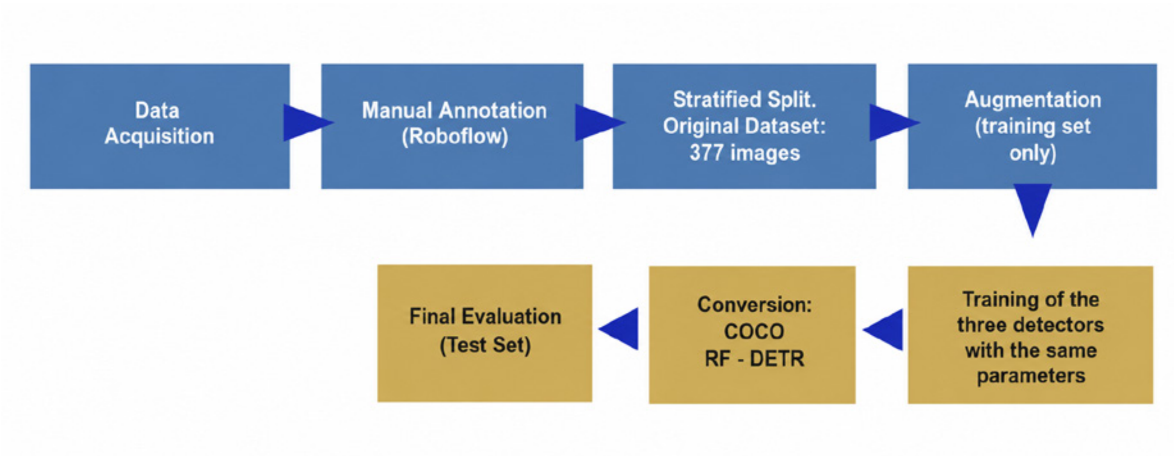

Figure 1

Complete experimental pipeline, from data acquisition to comparative evaluation

Note. The stages include manual annotation in Roboflow, stratified splitting of the original dataset (377 images), augmentation applied exclusively to the training set, conversion to COCO format for RF-DETR, training of the three detectors using the same baseline parameters (640 px, identical data splits), and final evaluation on the test set.

The dataset was compiled from multiple online sources and validated by the Bangladesh Agricultural Research Institute (BARI). It consists of 451 potato tuber images. In this study, 377 images were used after excluding the "miscellaneous" class to maintain a strictly phytosanitary focus. The six classes considered and their distribution were:

- Blackleg: 60 images
- Pink rot: 57
- Dry rot: 60
- Common scab: 62
- Black scurf: 58
- Healthy: 80

Object annotation and quality control

Manual lesion annotation was performed using Roboflow (Roboflow: Computer Vision Tools for Developers and Enterprises, 2025). Each lesion instance was defined by a bounding box, carefully adjusted to cover the visible lesion while minimizing the inclusion of adjacent healthy tissue. Annotation quality was controlled through the following procedures:

- Representative examples of each class were visually reviewed by a plant pathology specialist to verify annotation consistency.
- Bounding boxes that were excessively loose or did not correspond to the apparent lesion boundaries were corrected.
- Annotations were managed in both YOLO and COCO formats, generated from the same base annotation set to ensure compatibility with the training and evaluation frameworks.

Preprocessing

Before training, the following transformations were applied:

- Automatic orientation to normalize the rotation of the main object (tuber).
- Image resizing to 640 × 640 px while preserving the aspect ratio through letterboxing when necessary (Ultralytics default behavior).

Stratified data splitting

The original dataset of 377 images (before data augmentation) was stratified to preserve class proportions across all subsets:

- Training: 264 images (70%).
- Validation: 56 images (15%).
- Test: 57 images (15%).

Data augmentation (applied only to the training set)

To increase pattern diversity and mitigate overfitting, a data augmentation pipeline was applied exclusively to the 264 images in the training set. Transformations were performed in Roboflow while preserving geometric consistency between images and bounding boxes.

Image transformations:

- Horizontal flip.
- Fixed 90° rotations (clockwise and counterclockwise).
- Random rotation within the range [-15° to +15°].
- Random noise injection affecting up to 0.65% of image pixels.

Bounding box transformations:

- Horizontal flip synchronized with the image.
- Rotation within the range [-15° to +15°].
- Horizontal and vertical shear of ±10°.
- Limited positional noise (up to 0.16% of image pixels) to improve robustness to small annotation variations.

This process produced an augmented training set of 792 images (approximately a 3× augmentation factor). The validation (56 images) and test (57 images) sets remained unaugmented.

- Training: 792 images (~88%).
- Validation: 56 images (~6%).
- Test: 57 images (~6%).These percentages were calculated based on the total number of images used in the experiment (905 = 792 + 56 + 57).

The stratified split of approximately 88%/6%/6% was selected following common practices in deep-learning object detection tasks, where a large training set is prioritized to promote model convergence and generalization. Given the limited dataset size, a reduced but representative validation set was considered sufficient for monitoring training and tuning hyperparameters, while an independent test set was reserved for the final evaluation.

Stratification preserved class proportions across all subsets, reducing class imbalance bias and enabling a consistent comparison of model performance.

Detection models: YOLOv12 and YOLOv26

- YOLOv12 (nano) – yolov12n.pt. A recent detector focused on attention mechanisms to emphasize relevant features and improve the detection of small or partially occluded objects. The model was obtained from the official repository (Tian et al., 2025).
- YOLOv26 (nano) – yolov26n.pt. A variant optimized for edge devices with limited computational resources, prioritizing architectural simplicity and inference efficiency (Jocher & Qiu, 2026).
- RF-DETR (base). A real-time transformer architecture for object detection and instance segmentation developed by Roboflow (Robinson et al., 2026).

To ensure comparability, all models were trained and evaluated using the same dataset, identical train/validation/test partitions, the same input resolution (640 × 640 px), and the same training and evaluation protocol.

Software and hardware environment

The experiments were conducted in Google Colab using an NVIDIA Tesla T4 GPU (parameters shown in Figure 2). All software used was open-source and freely available: Ultralytics YOLO (AGPL-3.0 license), RF-DETR (Apache 2.0 license), and Google Colab (free version). Annotations were performed using Roboflow (free version for academic use).

Revista Científica CEDIA - ISSN: 3151-8435

```
+-----------------------------------------------------------------------------------------+
| NVIDIA-SMI 550.54.15              Driver Version: 550.54.15      CUDA Version: 12.4     |
|-----------------------------------------+------------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage | GPU-Util  Compute M. |
|                                         |                        |               MIG M. |
|=========================================+========================+======================|
|   0  Tesla T4                       Off |   00000000:00:04.0 Off |                    0 |
| N/A   35C    P8              9W /   70W |       0MiB /  15360MiB |      0%      Default |
|                                         |                        |                  N/A |
+-----------------------------------------+------------------------+----------------------+

+-----------------------------------------------------------------------------------------+
| Processes:                                                                              |
|  GPU   GI   CI        PID   Type   Process name                              GPU Memory |
|        ID   ID                                                               Usage      |
|=========================================================================================|
|  No running processes found                                                             |
+-----------------------------------------------------------------------------------------+
```

Figure 2

Google Colab training parameters

Training protocol

Initial weights: Pretrained weights were used for each corresponding “nano” variant (yolov12n.pt and yolov26n.pt) and the RF-DETR “base” model.

For the YOLO architecture (both models), the hyperparameters were:

- Epochs: 60
- Image size (imgsz): 640 px
- Batch size: 16
- Patience (early_stopping_patience): 10 (training stopped if the validation metric did not improve for 10 consecutive epochs)
- Configuration file (data.yaml): train/validation/test paths and six-class definition
- Data augmentation: as described in the “Stratified Data Splitting” section (applied consistently)

For the RF-DETR architecture, the hyperparameters were:

- Epochs: 15
- Image size (imgsz): 640 px
- Batch size: 4
- Patience (early_stopping_patience): 5

The differences in batch size and number of epochs reflect the characteristics of each architecture and memory constraints, which is common practice in comparative studies using different frameworks, while maintaining the dataset, input resolution, and evaluation criteria constant.

Evaluation and Metrics

The evaluation was performed exclusively on the test set (57 images). Standard Ultralytics metrics for object detection tasks were reported:

- mAP@0.5: mean average precision per class using an IoU threshold of 0.5.

$$(3)\ mAP@0.50 = \frac{1}{N}\sum_{i=1}^{N} AP_i(IoU \geq 0.50)$$

- mAP@0.5:0.95: mean of AP at multiple IoU thresholds [0.5: 0.95] with a step of 0.05, more demanding and sensitive to localization.

$$(4)\ mAP@0.50{:}\,0.95 = \frac{1}{10}\sum_{t=0.50}^{0.95} mAP_t$$

- F1-score: combines precision and recall.

$$(5)\ F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Results

Overall performance metrics

The three models were evaluated on the test set (n = 57 images) under identical conditions. Figure 3 shows the detections produced by each model, and Table 1 summarizes the overall results.

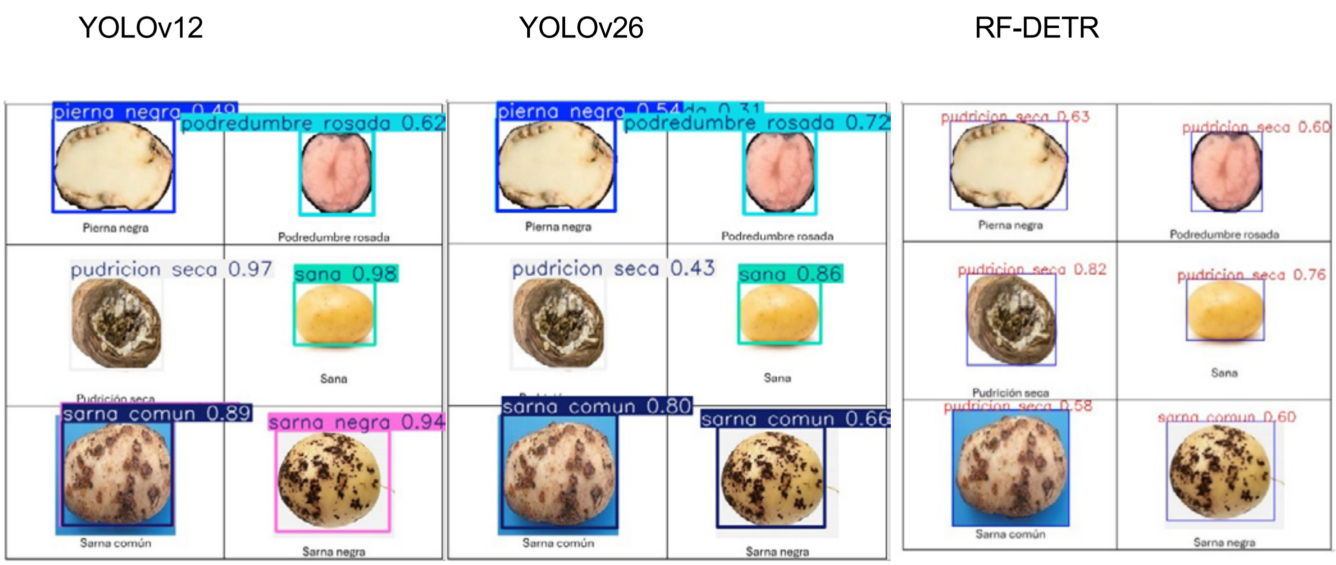

Figure 3

Visual comparison of the different models

Table 1
Overall results

Modelo	*Precision*	*Recall*	mAP50	mAP95
RF-DETR	70,86 %	88,39 %	85,61 %	82,64 %
YOLOv12	78,72 %	96,86 %	79,36 %	73,26 %
YOLOv26	79,34 %	96,64 %	78,48 %	73,55 %

Confusion matrix

Figure 4 (YOLOv12) and Figure 5 (YOLOv26) show the resulting confusion matrices; the diagonal axis displays the correct predictions for each class. In addition to the six study classes (blackleg, pink rot, dry rot, common scab, black scurf, and healthy), both matrices include a background row/column that aggregates:

- False positives (detections in regions without an assignable lesion).
- False negatives (missed lesions counted against the expected class as background).

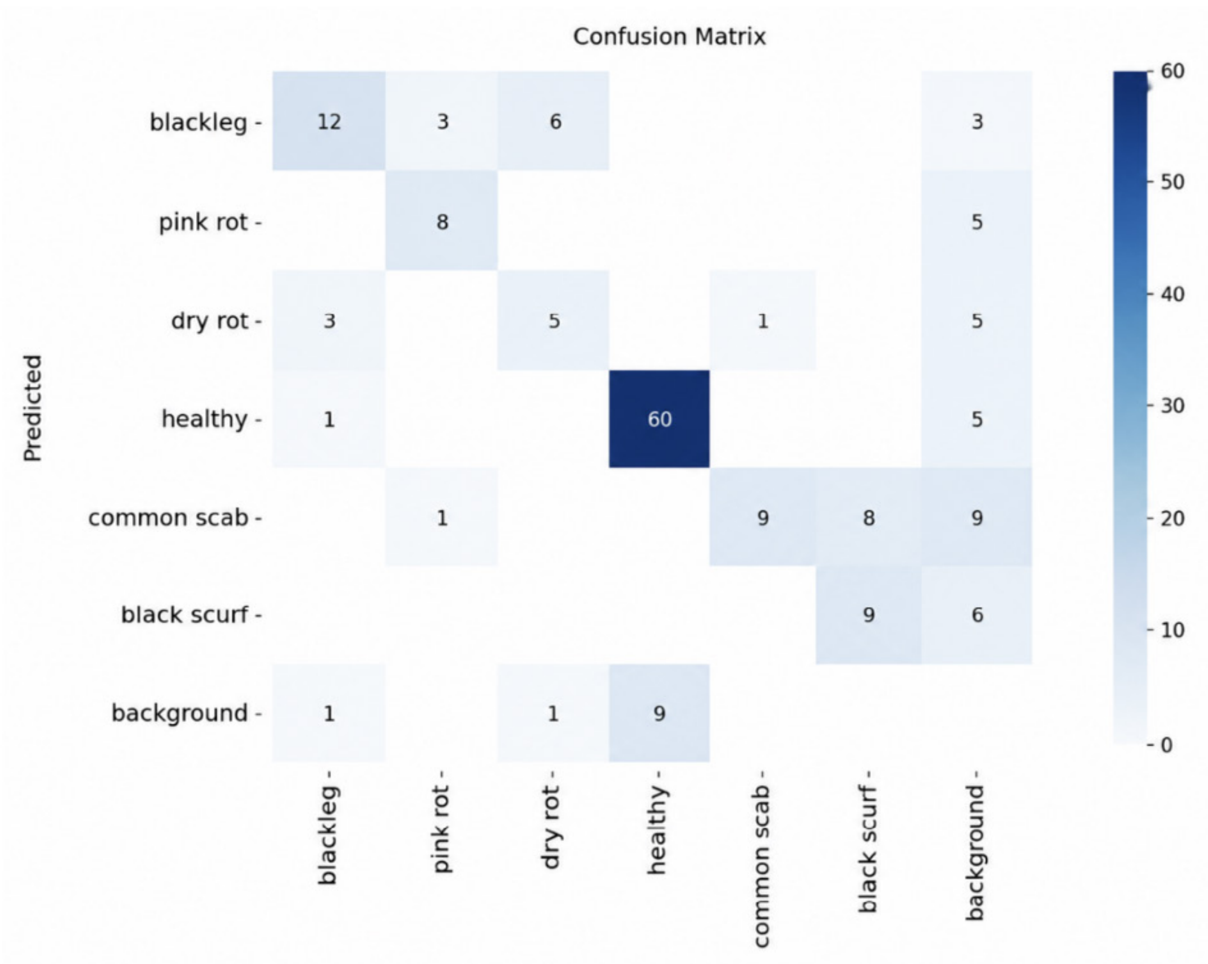

Figure 4

Confusion matrix – YOLOv12

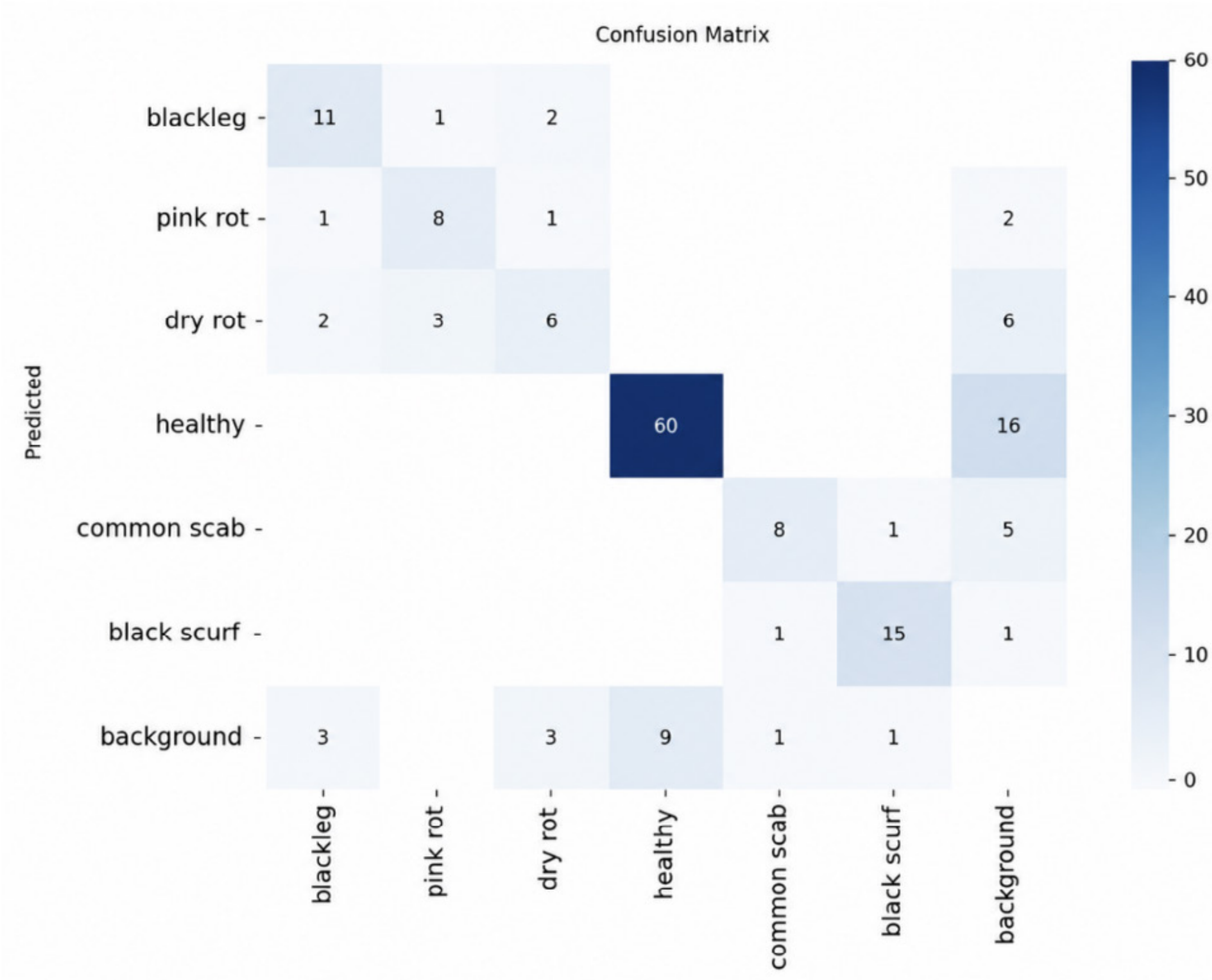

Figure 5

Confusion matrix – YOLOv26

Performance by class

Based on the confusion matrices, precision and recall were calculated for each class.

Revista Científica CEDIA - ISSN: 3151-8435

Table 2
Class-wise metrics – YOLOv12

Class	Precision	Recall
Blackleg	50,00 %	75,00 %
Pink rot	61,54 %	66,67 %
Dry rot	35,71 %	41,67 %
Common scab	33,33 %	90,00 %
Black scurf	60,00 %	52,94 %
Healthy	92,31 %	86,96 %

Revista Científica CEDIA - ISSN: 3151-8435

Table 3

Class-wise metrics – YOLOv26

Class	*Precision*	*Recall*
Blackleg	78,57 %	64,71 %
Pink rot	61,54 %	66,67 %
Dry rot	35,71 %	41,67 %
Common scab	33,33 %	90,00 %
Black scurf	60,00 %	52,94 %
Healthy	92,31 %	86,96 %

Revista Científica CEDIA - ISSN: 3151-8435

Table 4
Class-wise metrics – RF-DETR

Class	*Precision*	*Recall*
Blackleg	67,00 %	35,00 %
Pink rot	100,00 %	75,00 %
Dry rot	53,00 %	75,00 %
Common scab	50,00 %	100,00 %
Black scurf	93,00 %	82,00 %
Healthy	92,00 %	88,00 %

Key Findings: Comprehensive Comparison Between Models

Based on Tables 1, 2, 3, and 4, several important findings enable a comprehensive comparison between the models.

Overall comparison (accuracy and localization)

Based on the results (Table 1), the following observations can be made:

- *Overall precision (P): YOLOv26 achieved the highest value (79.34%), followed by YOLOv12 (78.72%) and RF-DETR (70.86%).*
- *Overall recall (R): YOLOv12 performed slightly better (96.86%), followed closely by YOLOv26 (96.64%) and above RF-DETR (88.39%).*
- *mAP@0.50: RF-DETR achieved the highest score (85.61%), outperforming YOLOv12 (79.36%) and YOLOv26 (78.48%).*
- *mAP@0.50:0.95 (more stringent): RF-DETR also ranked first (82.64%), while YOLOv26 (73.55%) and YOLOv12 (73.26%) achieved similar results.*

The YOLO models exhibited higher point precision and overall sensitivity at the default threshold (high capability to avoid missing lesions and to control false positives at a fixed operating point).

RF-DETR achieved higher Average Precision at both relaxed IoU (0.50) and strict IoU (0.50–0.95), suggesting that throughout the threshold sweep the model better preserves the area under the precision–recall curve and produces more consistent bounding boxes in terms of localization.

Practical implication: If the application requires maximum sensitivity and a low cost of missed detections (screening), either YOLO model is an appropriate option using the default threshold. If localization quality is the priority (e.g., severity estimation by lesion area, lesion cropping for segmentation, or integration into a downstream pipeline), RF-DETR offers a consistent advantage in mAP, particularly across the 0.50–0.95 range.

Class-wise comparison (precision, recall and F1)

Class-wise results

- Blackleg
- YOLOv26: P = 78.57% (highest), R = 64.71%, F1 ≈ 71.00% (best balance).
- YOLOv12: P = 50.00%, R = 75.00% (highest recall), F1 ≈ 60.00%.
- RF-DETR: P = 67.00%, R = 35.00%, F1 ≈ 46.00%.
- Conclusion: YOLOv26 provides the best balance for this class; YOLOv12 recovers more true positives (higher R), making it useful when minimizing missed cases is critical; RF-DETR requires threshold calibration or additional training samples.

Pink Rot

- RF-DETR: P = 100.00%, R = 75.00%, F1 ≈ 86.00% (best).
- YOLOv12/YOLOv26: P = 61.54%, R = 66.67%, F1 ≈ 64.00%.
- Conclusion: RF-DETR is superior, combining zero false positives with high sensitivity for this class.

Revista Científica CEDIA - ISSN: 3151-8435

Dry Rot

- RF-DETR: P = 53.00%, R = 75.00%, F1 ≈ 62.00% (best).
- YOLOv12/YOLOv26: P = 35.71%, R = 41.67%, F1 ≈ 39.00%.
- Conclusion: The morphological variability of this lesion (diffuse/subepidermal boundaries) negatively affects the YOLO models using bounding boxes; RF-DETR handles localization better and detects more cases.

Common scab

- RF-DETR: P = 50.00%, R = 100.00%, F1 ≈ 67.00% (best).
- YOLOv12/YOLOv26: P = 33.33%, R = 90.00%, F1 ≈ 48.00%.
- Conclusion: RF-DETR reduces the over-detection typically observed in the YOLO models (low P) while maintaining maximum recall.

Black scurf

- RF-DETR: P = 93.00%, R = 82.00%, F1 ≈ 88.00% (clearly superior).
- YOLOv12/YOLOv26: P = 60.00%, R = 53.00%, F1 ≈ 56.00%.
- Conclusion: RF-DETR better distinguishes the texture and tonal characteristics of this lesion, minimizing confusion with common scab and background.

Healthy

- RF-DETR: P = 92.00%, R = 88.00%, F1 ≈ 94.00%.
- YOLOv12/YOLOv26: P = 92.00%, R = 87.00%, F1 ≈ 89.00%.
- Conclusion: All three models are robust, with RF-DETR showing a slight advantage.

Revista Científica CEDIA - ISSN: 3151-8435

Table 5
Best model by class

Class	Best *precision*	Best *recall*	Best F1	Key comment
Blackleg	YOLOv26 (78.57 %)	YOLO v12 (75,00 %)	YOLOv26 (~71.00 %)	YOLOv26 provides better balance; YOLOv12 maximizes sensitivity.
Pink rot	RF-DETR **(100.00 %)**	RF-DETR (75,00 %)	RF-DETR (~86.00 %)	RF-DETR achieves zero false positives with high recall.
Dry rot	RF-DETR (53.00 %)	RF-DETR (75,00 %)	RF-DETR (~62.00 %)	RF-DETR handles diffuse/subepidermal boundaries better.
Common scab	RF-DETR (50.00 %)	RF-DETR (100.00 %)	RF-DETR (~67.00 %)	Reduces YOLO overdetection.
Black scurf	RF-DETR (93.00 %)	RF-DETR (82,00 %)	RF-DETR (~88.00 %)	Better distinguishes texture and dark tones.
Healthy	RF-DETR (92.00 %)	RF-DETR (88.00 %)	RF-DETR (~90.00 %)	Marginal differences; all models are robust.

Revista Científica CEDIA - ISSN: 3151-8435

Discussion

General overview and consistency with the literature

Under a common protocol (same data split, input resolution, and training protocol), the three detectors achieved competitive performance, with parity between YOLOv12 and YOLOv26 in the overall metrics (P/R and mAP) and an advantage in mAP for RF-DETR (both at IoU = 0.50 and across the 0.50–0.95 range). According to the results, YOLOv12 achieved precision = 78.72%, recall = 96.86%, mAP@0.50 = 79.36%, and mAP@0.50:0.95 = 73.26%, whereas YOLOv26 obtained precision = 79.34%, recall = 96.64%, mAP@0.50 = 78.48%, and mAP@0.50:0.95 = 73.55%. In numerous contemporary agricultural applications, the YOLO family has become predominant because of its balance between speed and accuracy and its practical adoption in the field (including tubers and postharvest tasks), which is consistent with recent systematic reviews and benchmarking studies (Badgujar et al., 2024; Divyanth et al., 2024).

The DETR family—and particularly its real-time variants—eliminates the need for NMS and leverages query-based bipartite matching, which often results in better localization consistency at strict IoU thresholds and precision–recall curves with greater area (AP) across a wide range of thresholds (Dahiya et al., 2025). This behavior has been documented for RT-DETR (CVPR 2024) (Zhao et al., 2024) and its implementations, which demonstrate simultaneous improvements in speed and AP compared with one-stage detectors and other classical DETR architectures. In addition, agricultural applications using RF-DETR have begun to appear, reinforcing the relevance of including a transformer-based detector in this comparison (Blok et al., 2025; Öztürk & Şin, 2026).

Overall metrics: Point P/R vs. Integrated AP (IoU = 0.50 and 0.50–0.95)

Point metrics (P and R) depend on the confidence threshold and post-processing (e.g., NMS in YOLO), whereas mAP@0.50 and mAP@0.50:0.95 integrate the area under the precision–recall curve across threshold sweeps (and, in COCO, across multiple IoU thresholds), providing a threshold-independent measure of classification and localization. This difference explains why a model may exhibit superior AP while simultaneously showing lower P/R at a given operating point: the global optimum of the curve does not necessarily coincide with the threshold used during validation or inference (Thibaut, 2023).

In the present results, RF-DETR achieved higher mAP at both 0.50 and 0.50–0.95, consistent with the literature reporting improved localization at strict IoU thresholds, elimination of NMS, and larger precision–recall areas due to higher-quality queries and efficient decoders (Parlak, 2026; Rahaman Sams et al., 2026). In contrast, YOLOv12 and YOLOv26 retained slightly better point recall and precision at the default Ultralytics threshold, a common pattern in one-stage detectors with NMS and highly optimized calibration for agricultural applications (Badgujar et al., 2024; Divyanth et al., 2024).

Class-wise performance: comparative interpretation and probable causes

Blackleg. YOLOv26 achieved the highest F1-score (P ≈ 79.00%; R ≈ 65.00%), providing the best balance between precision and sensitivity, whereas YOLOv12 retained the highest recall (75.00%), making it suitable for screening. The lower recall of RF-DETR suggests a suboptimal threshold or the need for additional samples for this specific class. This type of precision–recall trade-off between architectures has been observed in agricultural benchmarks, where texture and moisture influence inter-class confusion (Badgujar et al., 2024).

Pink Rot. RF-DETR performed best (P = 100.00%; R = 75.00%), achieving zero false positives and high recall at the selected operating point. The precision–recall curve for this class tends to dominate across thresholds, consistent with the advantages of transformers in contextual representation and spatial consistency (Zhao et al., 2024).

Dry Rot. RF-DETR outperformed YOLO in recall and F1-score. Its superior performance with diffuse lesion boundaries and subepidermal manifestations suggests that the cost function (L1 + GIoU) and one-to-one matching reduce the penalty at high IoU thresholds. Nevertheless, the literature agrees that internal or incipient lesions require NIR/SW-NIR/HSI imaging to improve sensitivity, reinforcing a future multimodal approach (Prasetyo et al., 2024; F. Zhang et al., 2024).

Common Scab vs. Black Scurf. The YOLO models exhibited over-detection of common scab (low precision, high recall), whereas RF-DETR improved the balance (higher precision and recall), probably because it better captures the granular texture and spatial context of Rhizoctonia solani sclerotia (black scurf) and their variability (AG-3, etc.), which are known to produce visual ambiguities in potato tubers (Chaudhary et al., 2023; Potato - Rhizoctonia Solani (Stem Canker & Black Scurf), 2025).

Healthy. All three models were robust (high precision and recall, with a slight advantage for RF-DETR), a result consistent with applications involving tubers and postharvest inspection as well as reviews of YOLO-based agricultural applications (Badgujar et al., 2024; Divyanth et al., 2024).

Statistical validity and transparency

Given the size of the test set (n = 57), small differences (≤1 percentage point) may not be statistically significant. It is recommended to apply image- or instance-level bootstrapping to obtain 95.00% confidence intervals for precision, recall, mAP, and F1-score, as well as permutation tests to evaluate the null hypothesis of “no difference.” In addition, reporting precision–recall curves together with the points of maximum F1 (overall and per class) is recommended to align the operating point with the desired risk policy (cost of false negatives versus false positives). These practices are consistent with the COCO evaluation protocol and recent guidelines for mAP and precision–recall analysis (Thibaut, 2023; F. Zhang et al., 2024).

Practical ICT implications

- 1. System architecture (edge ↔ cloud)

The results support a hybrid design consisting of edge inference (cameras/SoC/Jetson) for real-time response and event-driven transmission of ambiguous cases to the cloud for auditing and re-annotation. The YOLO family is widely adopted in agriculture because of its low latency and ease of deployment, whereas

RF-DETR provides end-to-end detection without NMS, simplifying the post-processing pipeline on resource-constrained devices while offering superior AP at strict IoU thresholds (Badgujar et al., 2024; F. Zhang et al., 2024).

2. Model selection according to the use case (operating points)

For high-throughput screening (avoiding missed lesions), a YOLO model calibrated with a low confidence threshold maximizes recall. For diagnosis requiring severity quantification (high IoU), RF-DETR provides higher mAP@0.50:0.95 and more consistent bounding boxes throughout the threshold sweep. Precision–recall curves and COCO mAP are the recommended tools for selecting class-specific thresholds according to the cost of errors (Thibaut, 2023; Zhao et al., 2024).

- 3. *Two-Stage Pipelines (Screening → Refinement)*

A YOLO → RF-DETR workflow reduces computational cost on edge devices (fast detection) while increasing precision and IoU for positive detections (NMS-free refinement), a strategy already suggested in agricultural computer vision benchmarking and consistent with the real-time RT-DETR evidence (Zhou et al., 2025).

- 4. Multimodality for internal lesion

For dry rot and other subepidermal lesions, integrating NIR/SW-NIR/HSI into the ICT pipeline improves early detection, overcoming the limitations of RGB imaging. Late fusion or routing suspicious samples to an NIR/HSI inspection station are scalable strategies for processing plants operating under cycle-time constraints (Prasetyo et al., 2024; F. Zhang et al., 2024).

Conclusions

1. Comparative Performance and Operational Feasibility

Under the same experimental protocol—identical dataset, stratified split, and input resolution (640 px)—the YOLO family detectors and the transformer-based RF-DETR model proved to be viable for the early detection of diseases in postharvest potato tubers. YOLOv12 and YOLOv26 showed comparable performance in terms of precision and recall, whereas RF-DETR achieved higher mAP values at both IoU = 0.50 and across the 0.50–0.95 range. These results indicate the suitability of both paradigms for integration into automated agricultural inspection systems with moderate computational resources.

2. Sensitivity, localization, and selection according to the use case

The comparative analysis revealed a systematic trade-off between sensitivity and localization accuracy. The YOLO models exhibited slightly higher recall at the default threshold, making them suitable for large-scale screening scenarios where minimizing false negatives is a priority. In contrast, RF-DETR demonstrated stricter and more consistent localization, reflected in its advantage in mAP@0.50:0.95, attributed to its one-to-one assignment mechanism and the absence of non-maximum suppression. This behavior is particularly relevant when precise lesion delineation is required for severity estimation or subsequent classification.

3. Class-wise Behavior and Limitations of the RGB Approach

The confusion matrix analysis identified error patterns specific to the agricultural domain. In particular, the confusion between common scab and black scurf is explained by phenotypic similarities in surface texture and coloration, as reported in the phytopathological literature. Likewise, subepidermal lesions associated with Fusarium spp. exhibited higher false-negative rates and lower spatial overlap, highlighting an inherent limitation of using RGB images alone for the early detection of internal damage.

4. Practical implications and recommended architecture

Based on the obtained results, a two-stage functional architecture is proposed: an initial YOLO-based detection stage aimed at maximizing sensitivity and enabling rapid screening, followed by a second stage using RF-DETR for localization refinement and IoU improvement. This combination balances computational efficiency and spatial precision and can be further enhanced through calibration techniques and, when the objective is to reduce missed detections of incipient lesions, by late fusion with NIR or HSI spectral information. Overall, the findings suggest the feasibility of developing an automated, accurate, and scalable tool aligned with the requirements of real-time monitoring systems and edge computing in postharvest environments.

Conflict of interest statement

The authors declare that they have no conflicts of interest.

Funding sources

The author declares that this research received no external funding from any public, commercial, or non-profit funding agency.

Author contributions (CRediT):

Manolo Sebastián Muñoz-Espinoza: Conceptualization; methodology; software; validation; formal analysis; investigation; resources; data curation; writing – original draft; writing – review and editing; visualization; supervision; project administration.

Referencias

Aleinzi, M. (2025). Enhancing precision agriculture with YOLOv8: A deep learning approach to potato disease identification. I*nternational Journal of Advanced Computer Science and Applications,* 16(4), 1156–1166.
https://doi.org/10.14569/IJACSA.2025.01604110

Badgujar, C. M., Poulose, A., & Gan, H. (2024). Agricultural object detection with You Look Only Once (YOLO) algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture, 223*.
https://doi.org/10.1016/j.compag.2024.109090

Blok, P. M., Wang, H., Suh, H. K., Wang, P., Burridge, J., & Guo, W. (2025). PointRAFT: 3D deep learning for high-throughput prediction of potato tuber weight from partial point clouds [Preprint]. *arXiv.*
https://doi.org/10.48550/arXiv.2512.24193

Chaudhary, S., Lal, M., Sagar, S., Sharma, S., & Kumar, M. (2023). Black scurf of potato: Insights into biology, diagnosis, detection, host-pathogen interaction, and management strategies. *Tropical Plant Pathology, 49*(2), 169–192.
https://doi.org/10.1007/s40858-023-00622-4

Dahiya, N., Prakash, D., Kundu, S., Kuttan, S. R., Suwalka, I., Ayadi, M., Dubale, M., & Hashmi, A. (2025). Optimised RFO tuned RF-DETR model for precision urine microscopy for renal and systemic disease diagnosis. *Scientific Reports, 15*(1), 25842.
https://doi.org/10.1038/s41598-025-11725-0

Divyanth, L. G., Khanal, S. R., Paudel, A., Mattupalli, C., & Karkee, M. (2024). Efficient detection of eyes on potato tubers using deep-learning for robotic high-throughput sampling. Frontiers in *Plant Science, 15,* 1512632.
https://doi.org/10.3389/fpls.2024.1512632

Ephytia. (2025, July 24). Potato—Rhizoctonia solani (stem canker & black scurf). https://ephytia.inra.fr/en/C/20899/Potato-Rhizoctonia-solani-Stem-canker-Black-scurf

Faria, F. T. J., Moin, M. Bin, & Wase, A. Al. (2023). *Potato Diseases Datasets* [Conjunto de datos]. Kaggle.
https://doi.org/10.34740/KAGGLE/DSV/7203264

Geoffrey Manurung, D., Ryan Pinasthika, M., Obila Vasya, M. A., Dwi Setya Putri, R. A., Parasian Tampubolon, A., Fadhil Prayata, R., Khoirin Nisa, S., & Yudistira, N. (2024). Deteksi dan klasifikasi hama potato beetle pada tanaman kentang menggunakan YOLOV8. *Jurnal Teknologi Informasi dan Ilmu Komputer, 11*(4), 723–734. https://doi.org/10.25126/jtiik.1148092

Guillemette, A. M., Casanova, G. H., Hamilton, J. P., Pokorná, E., Dobrev, P. I., Motyka, V., Rashotte, A. M., & Leisner, C. P. (2024). The physiological and molecular responses of potato tuberization to projected future elevated temperatures. *Plant Physiology, 197*(1), 664. https://doi.org/10.1093/PLPHYS/KIAE664

He, Y., Peng, Y., Wei, C., Zheng, Y., Yang, C., & Zou, T. (2024). Automatic disease detection from strawberry leaf based on improved YOLOv8. Plants, 13(18). 2256 https://doi.org/10.3390/PLANTS13182556

Jayanthi, J., & Kumar, K. A. (2024). Transformative impact of AI-driven computer vision in agriculture. En *Artificial intelligence techniques in smart agriculture* (pp. 129–150). Springer https://doi.org/10.1007/978-981-97-5878-4_9

Jocher, G., & Qiu, J. (2026). *Ultralytics YOLO26* [Software]. GitHub. https://github.com/ultralytics/ultralytics

Keithellakpam, L. B., Karunakaran, C., Singh, C. B., Jayas, D. S., & Danielski, R. (2026). A comprehensive review on pre- and post-harvest perspectives of potato quality and non-destructive assessment approaches. *Applied Sciences, 16*(1). 190 https://doi.org/10.3390/APP16010190

Liu, C. (2024). Advancing strawberry disease detection in agriculture: A transfer learning approach with YOLOv5 algorithm. *International Journal of Advanced Computer Science and Applications, 15*(3), 1013–1022. https://dx.doi.org/10.14569/IJACSA.2024.01503101

Ma, P., Li, C., Rahaman, M. M., Yao, Y., Zhang, J., Zou, S., Zhao, X., & Grzegorzek, M. (2022). A state-of-the-art survey of object detection techniques in microorganism image analysis: From classical methods to deep learning approaches. *Artificial Intelligence Review, 56*(2), 1627–1698. https://doi.org/10.1007/S10462-022-10209-1

Muñoz Espinoza, M., Moreno Castillo, W. E., y Palacios Ruiz, F. P. (2026). Detección de enfermedades en papa mediante deep learning usando YOLOv12. Ciencias de la Ingeniería y *Aplicadas, 10*(1), 119–129. https://doi.org/10.61236/CIYA.V10I1.1239

Moreno Mayhuire, J. S., & Herrera Quispe, J. A. (2023). Modelos de aprendizaje automático para la clasificación de enfermedades de la *Solanum tuberosum*: Una revisión sistemática de la literatura. *Revista de Investigación de Sistemas e Informática, 16*(2), 129–137. https://doi.org/10.15381/risi.v16i2.25980

Nagel, M., Dulloo, M. E., Bissessur, P., Gavrilenko, T., Bamberg, J., Ellis, D., & Giovannini, P. (2022). *Global strategy for the conservation of potato*. Leibniz Institute of Plant Genetics and Crop Plant Research. https://doi.org/10.5447/IPK/2022/29

Öztürk, L., & Şin, B. (2026). Real-time transformer-based object detection for advanced plant disease recognition: The grapevine fanleaf virus example. *Applied Fruit Science, 68*(1). https://doi.org/10.1007/s10341-026-01777-5

Palacio Betancur, S., & Bolaños Martínez, F. (2025). Tomato (*Solanum lycopersicum L.*) leaf disease detection using computer vision. *Revista Facultad Nacional de Agronomía, 78*(3), 11203–11212. https://doi.org/10.15446/rfnam.v78n3.116493

Parlak, C. (2026). Evaluation of cutting-edge object detection architectures on multi-object and single-object datasets. *Black Sea Journal of Engineering and Science*, (9)1. 287-294. https://doi.org/10.34248/bsengineering.1736319

Prasetyo, E. W., Amanah, H. Z., Farras, I., Pahlawan, M. F. R., & Masithoh, R. E. (2024). Detection of *Fusarium spp*. infection in potato (*Solanum tuberosum L.*) during postharvest storage through visible-near-infrared and shortwave-near-infrared reflectance spectroscopy. *Open Agriculture, 9*(1). https://doi.org/10.1515/opag-2022-0295

Qin, R., Wang, Y., Liu, X., & Yu, H. (2024). Advancing precision agriculture with deep learning enhanced SIS-YOLOv8 for Solanaceae crop monitoring. *Frontiers in Plant Science, 15,* 1485903. https://doi.org/10.3389/fpls.2024.1485903

Rahaman Sams, F., Kazi Supti, S., Binte Hamid, S., Junayed, R., Fahim A Bari, K. M., Junaeid Ali, M., Gani, R., Shams, K., Rifat Ahmmad Rashid, M., & Ul Islam, R. (2026). Real-time sunflower detection using semi-supervised and self-supervised deep learning for precision agriculture. *Smart Agricultural Technology*, *13*, 101684. https://doi.org/10.1016/J.ATECH.2025.101684

Redmon, J. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE *Conference on Computer Vision and Pattern Recognition.*

Robinson, I., Robicheaux, P., Popov, M., Ramanan, D., & Peri, N. (2026). RF-DETR: Neural architecture search for real-time detection transformers [Preprint]. arXiv. https://arxiv.org/abs/2511.09554

Roboflow. (2025). *Roboflow: Computer vision tools for developers and enterprises*. https://roboflow.com/

Sapkota, R., Cheppally, R. H., Sharda, A., & Karkee, M. (2025). YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection [Preprint]. arXiv. https://arxiv.org/pdf/2509.25164

Sharma, S., Tiwari, R. K., Sagar, V., & Maharana, C. (2024). Soil- and tuber-borne diseases of potato. En *Approaches for potato crop improvement and stress management* (pp. 179–231). https://doi.org/10.1007/978-981-97-1223-6_7

Solankey, S. S. (Ed.). (2025). *Advances in research on potato production*. Springer. https://doi.org/10.1007/978-3-031-82710-5

Thibaut, L. (2023, August 13). *COCO evaluation metrics explained.* Picsellia. https://www.picsellia.com/post/coco-evaluation-metrics-explained

Tian, Y., Ye, Q., & Doermann, D. (2025, February 20). *YOLO12: Attention-centric real-time object detectors* [Preprint]. arXiv. https://arxiv.org/abs/2502.12524

Upeniece, L., Bimšteine, G., & Stramkale, V. (2025). The most important potato diseases and microbiological agents for their control: A review. Research for Rural Development 2025: *Annual 31st International Scientific Conference Proceedings* (Vol. 1, pp. 17–23). Latvia University of Life Sciences and Technologies. https://doi.org/10.22616/RRD.31.2025.003

Yurdakul, M., Sazak, H., Kotan, M., & Ta¸sdemir, T. (2026). *A review of YOLO family from YOLOv1 to YOLO26* [Preprint]. Preprints.org https://doi.org/10.20944/PREPRINTS202602.1844.V1

Zhang, F., Wang, W. X., Wang, C. S., Zhou, J., Pan, Y., & Sun, J. F. (2024). Study on hyperspectral detection of potato dry rot in gley stage based on convolutional neural network. Spectroscopy and *Spectral Analysis*, 44(2), 480-489. https://doi.org/10.3964/j.issn.1000-0593(2024)02-0480-10

Zhang, J., Xie, J., Zhang, F., Gao, J., Yang, C., Song, C., Rao, W., & Zhang, Y. (2024). Greenhouse tomato detection and pose classification algorithm based on improved YOLOv5. *Computers and Electronics in Agriculture*, 216, 108519. https://doi.org/10.1016/J.COMPAG.2023.108519

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., & Chen, J. (2024). DETRs beat YOLOs on real-time object detection. En Proceedings of the IEEE/CVF *Conference on Computer Vision and Pattern Recognition* (pp. 16965–16974). https://zhao-yian.github.io/RTDETR.

Zhou, H., Liu, Y., & Wang, J. (2025). Railway fastener defect detection using RFD-DETR: A lightweight real-time transformer-based approach. *PLOS ONE*, 20(11), e0331513. https://doi.org/10.1371/JOURNAL.PONE.0331513