

Deep-learning-based Detection of Citrus Pests and Diseases via Image Recognition

Ping Wang,^{1,2*} Mideth Abisado,¹ Yu Chen,^{1,3} and Xianwei Zeng^{2**}

¹College of Computing and Information Technologies, National University, Manila, Philippines

²College of Artificial Intelligence, Yango University, Mawei District, Fujian 350015, China

³College of Electronic Engineering, Fuzhou Institute of Technology, Fuzhou, Fujian 350506, China

(Received August 31, 2025; accepted October 16, 2025)

Keywords: citrus disease detection, deep learning, YOLO, precision agriculture, edge computing

Crop production in subtropical regions such as Fujian, China, is highly vulnerable to the rapid spread of pests and diseases due to warm and humid climatic conditions. In this study, we propose a lightweight, deep-learning-based detection system tailored for citrus disease diagnosis, focusing on three categories: Huanglongbing (HLB), citrus canker, and healthy leaves. A curated image dataset was constructed and used to train several object detection models, with You Only Look Once (YOLO) v3 and YOLOv4 variants showing exceptional performance. The best-performing model, YOLOv3 in its fine-tuned Phase 2 version, achieved average precision scores of 98.6% for HLB, 97.4% for citrus canker, and 98.7% for healthy leaves. These results validate the system's ability to accurately distinguish disease states under field conditions. The proposed framework supports early-stage detection, significantly reduces labor burden, and is optimized for deployment on edge devices, enabling real-time monitoring in agricultural environments. This work demonstrates a scalable and efficient solution for intelligent citrus crop management.

1. Introduction

According to global demographic assessments, the world population continues to grow, which will intensify the demand pressure on Earth's resources and lead to increasingly severe challenges to global food security. In regions highly dependent on imported food, food self-sufficiency remains relatively low, underscoring their vulnerability in the face of a global food crisis. Additionally, extreme weather events induced by climate change have significantly increased the uncertainty of agricultural production, resulting in price surges and supply shortages. Against this background, developing efficient and stable crop production methods and enhancing yield have become urgent priorities.

Precision agriculture, as a modern agricultural management approach integrating advanced technologies, aims to scientifically manage crop growth through information technology and big data analysis, thus improving the productivity and efficiency of agricultural production.⁽¹⁾ Crop

*Corresponding author: e-mail: wangp@students.national-u.edu.ph

**Corresponding author: e-mail: 507251393@qq.com

<https://doi.org/10.18494/SAM5916>

disease identification technology, as a key aspect of precision agriculture, enables the timely monitoring of disease conditions, preventing the spread of diseases and effectively improving crop health and yield.

In recent years, the aging population and declining birthrate have exacerbated the reduction in agricultural labor, making the traditional manual detection of pests and diseases time-consuming, labor-intensive, and susceptible to subjective biases. To address this, in this study, we attempt to apply single-stage object detection network technology, combined with deep learning algorithms, to design and implement a system for detecting diseases in citrus leaves, utilizing edge computing to automate disease detection.^(2,3) Single-stage object detection networks have been widely applied in industrial and smart life domains, and in recent years, they have gradually penetrated the agricultural field, becoming an important tool for automated plant disease recognition.

Edge computing, as a form of distributed computing architecture, allows data processing to occur closer to IoT devices or local edge servers, facilitating rapid access to analysis and prediction results. In agriculture, edge computing technology can be integrated with environmental monitoring systems installed in fields or greenhouses to collect real-time information on sunlight, temperature, humidity, and soil moisture, providing data support for crop health assessment and pest and disease early warning. The ultimate goal of this research is to develop a low-cost, feasible, and efficient citrus disease recognition system as an effective solution for smart agriculture.

2. Literature Review

Traditional methods for plant disease detection primarily rely on manual identification, but these methods are time-consuming, low in accuracy, and susceptible to environmental interference. Hyperspectral imaging technology is one of the common methods for detecting plant diseases, capturing more spectral information to distinguish between healthy and diseased plants.⁽⁴⁾ However, hyperspectral imaging equipment is expensive, limiting its accessibility in agricultural production. To address this issue, academia has increasingly turned to machine learning and deep learning technologies, enabling automated diagnosis by analyzing disease characteristics on plant leaves.

Machine learning offers certain advantages in disease detection by classifying on the basis of the color, contour, and other features of images. However, traditional machine learning methods are sensitive to background interference, making it challenging to achieve stable detection results in complex backgrounds. In contrast, convolutional neural networks (CNNs) demonstrate strong adaptability in image feature extraction, allowing them to ignore background interference, identify key features of plant diseases, and significantly improve detection accuracy.^(5–11)

As a significant global economic crop, citrus is widely cultivated in tropical and subtropical regions. Major citrus diseases include Huanglongbing (HLB) (citrus greening) and citrus canker.^(12,13)

- HLB (citrus greening): HLB is a systemic disease characterized by yellowing veins, the regeneration of small and curled leaves, and early wilting and shedding. In severe cases, the

tree quickly declines, leading to its death. This disease spreads through grafting infected with pathogenic bacteria, severely impacting the economic value of citrus trees.

- Citrus canker: Citrus canker mainly affects leaves, twigs, and fruits, presenting as small, water-soaked, dark green spots that gradually develop into spongy, corky lesions. The lesions contain a high concentration of pathogenic bacteria, and the disease spreads easily in rainy climates, adversely affecting tree health and yield.

3. Methods

The model developed to detect citrus diseases was evaluated for the health status of citrus leaves using deep learning. We utilized CNN to extract features from both healthy and diseased citrus leaves, distinguishing between them on the basis of these features. Adjustments were made to the You Only Look Once (YOLO)-based model architecture to enable the designed model to operate on embedded devices.

YOLOv3 and YOLOv4 are preferred in plant disease diagnosis owing to their high accuracy, speed, and efficiency. YOLOv4, for instance, achieves near-perfect results on datasets such as Plant Village with minimal latency. Both models are lightweight, so they are appropriate for real-time detection on mobile or edge devices, which is crucial in agriculture. They can process imbalanced data and deal with symptom variability effectively. Newer YOLO versions require more computational resources and tuning without offering significant benefits for this task. Proven reliability and strong community support the use of YOLOv3 or v4 in deployment.

Figure 1 shows the citrus disease detection system, in which the input image is processed through the initial object detection network designed for this study. The network extracts features from the image and achieves disease classification and localization by fusing feature maps from different levels.

3.1 Citrus disease database

The dataset for this study was sourced from publicly available agricultural databases, along with images of citrus infections from local agricultural research stations, totaling nearly 900

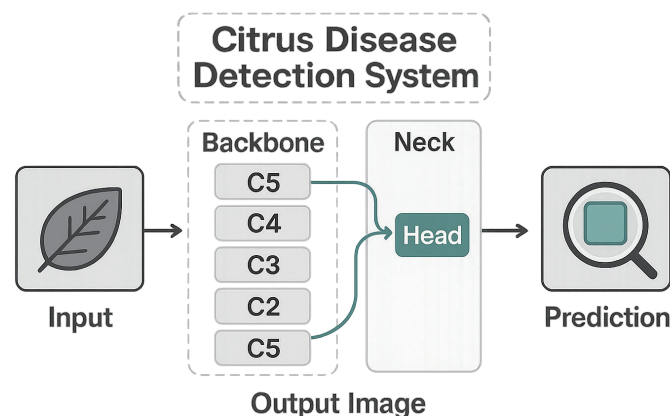


Fig. 1. (Color online) Citrus disease detection system.

images. The training data are categorized into three classes: HLB, canker disease, and healthy leaves. Figure 2 displays (a) a leaf affected by HLB, (b) a leaf with canker disease, and (c) a healthy leaf.

Before training the object detection model, images were processed by annotating the coordinates and categories of objects within each image. LabelImg software was used to annotate object locations, generating eXtensible Markup Language files in a format containing the necessary object location and category information for training. To expand the dataset, data augmentation techniques were applied, such as rotation and blurring, generating additional training data. This process increased the original dataset to a few thousand images to reduce the risk of overfitting due to insufficient data.

The image dataset was expanded to improve the performance of the models. This process involves data augmentation techniques to artificially increase the dataset size, which include rotation, flipping, scaling, brightness adjustment, and adding noise. By using Roboflow and OpenCV as automated tools, the augmentation process was completed in a few hours. After augmentation, each image was annotated with bounding boxes and class labels. Dataset expansion accounted for 30–50% of the total process time. Once the dataset was prepared, training and inference using the models became fast.

3.2 Model design

YOLOv4 is a one-stage object detection network. We designed two lightweight variants on the basis of the YOLOv4 architecture, optimized for edge deployment. To reduce computational load while maintaining detection accuracy, the model input size is fixed at $416 \times 416 \times 3$ (RGB). The backbone network is compressed by reducing the number of layers, and PANet is replaced with feature pyramid networks (FPNs) for feature fusion. Figures 3 and 4 illustrate the network architectures of Models 1 and 2, respectively. The difference between the two lies in the incorporation of squeeze-and-excitation attention blocks in the final two convolutional layers of Model 2, which dynamically reweights feature maps to enhance classification performance.

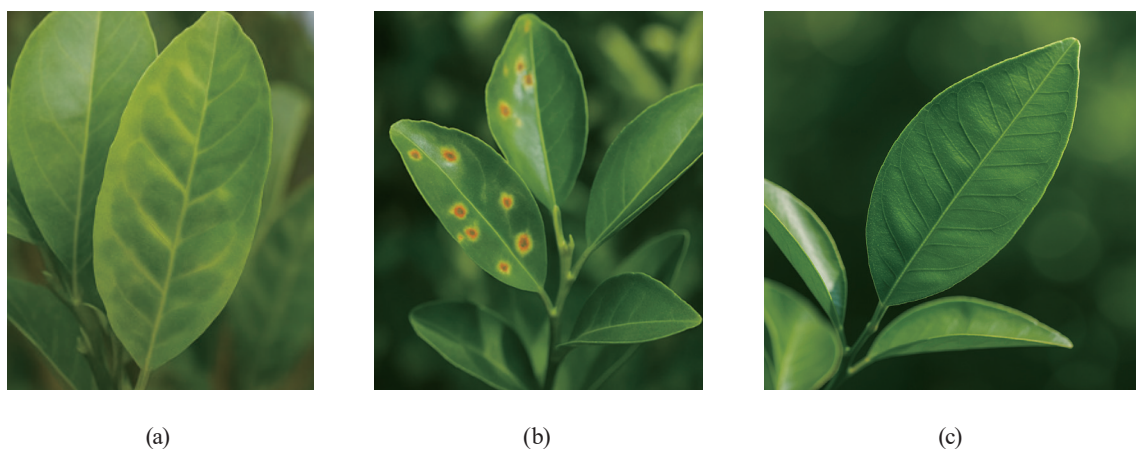


Fig. 2. (Color online) Images used for model training: (a) HLB leaf, (b) canker disease leaf, and (c) healthy leaf.

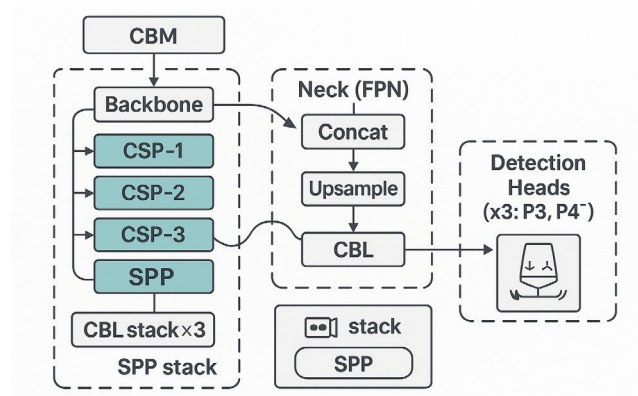


Fig. 3. (Color online) Object detection model 1 designed in this study.

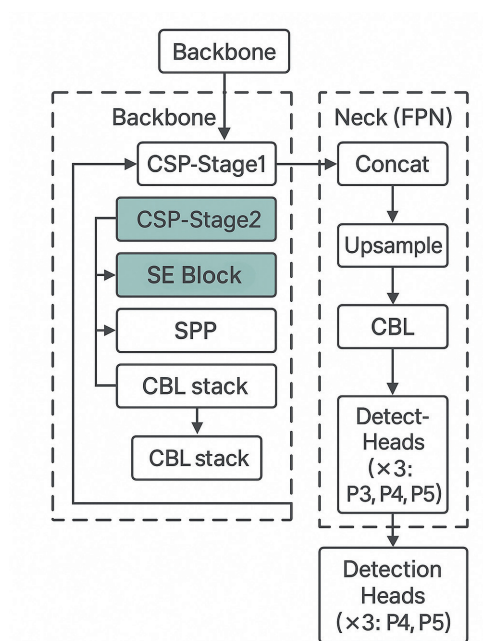


Fig. 4. (Color online) Object detection model 2 designed in this study.

3.3 Model training

The training process of the object detection model comprises three main stages: data preparation, model training, and performance testing. Initially, original images were augmented using techniques such as rotation, scaling, brightness adjustment, and flipping. Each image was then annotated using LabelImg, marking disease-affected areas with bounding boxes. The dataset was split in an 8:2 ratio into training and test sets for learning and validation, respectively.

The training phase employed a YOLOv4-based architecture, incorporating Batch Normalization and the Mish activation function. Backpropagation was used to iteratively update

the network weights for the optimized feature learning of citrus leaf diseases. The initial learning rate was set to 0.001 and adjusted dynamically by cosine annealing. The batch size was 64, and the training was conducted for 500 epochs. To prevent overfitting, dropout layers and early stopping criteria were applied. After training, the model was evaluated using the test set, with metrics including precision, recall, and mean average precision (mAP) to assess its generalization capability.

3.4 Object detection network evaluation

To assess the performance of the detection network, we used mAP. The model's classification behavior was analyzed using a confusion matrix (Table 1).

Precision and recall are calculated using the following equations:

$$Precision = TP/(TP + FP), \quad (1)$$

$$Recall = TP/(TP + FN). \quad (2)$$

For object detection, bounding box predictions are evaluated using the Intersection over Union (IoU) metric. If the IoU metric between a predicted box and the ground truth exceeds a threshold (typically 0.5), it is counted as *TP*; otherwise, it is considered *FP*.

To visualize detection performance, a precision–recall (PR) curve was plotted with recall on the *x*-axis and precision on the *y*-axis (Fig. 5). The area under the PR curve, referred to as

Table 1
Confusion matrix.

	Actual positive	Actual negative
Predicted positive	True positive (<i>TP</i>)	False positive (<i>FP</i>)
Predicted negative	False negative (<i>FN</i>)	True negative (<i>TN</i>)

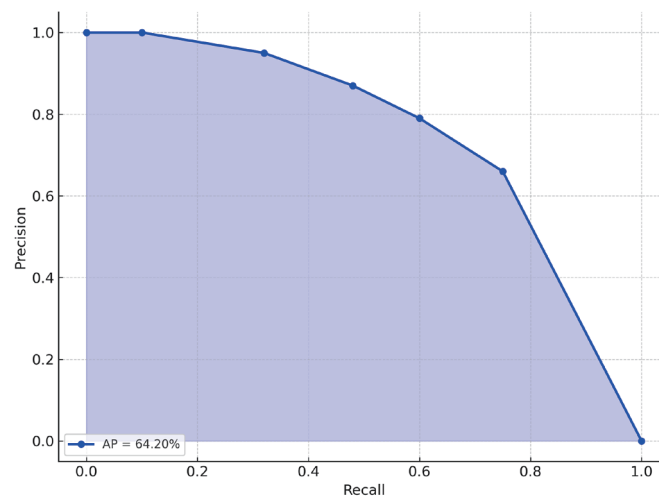


Fig. 5. (Color online) PR curve.

average precision (AP), was adopted as a key performance indicator. The PR curve illustrates the model's predictive performance on the test dataset. The calculated AP value is 78.42%,.

4. Results and Discussion

The performance characteristics of the four object detection models, Model 1, Model 2, YOLOv3, and YOLOv4, were evaluated and compared in two phases. Each model's performance was assessed using mAP and per-class AP. The models were trained to identify three citrus leaf categories: HLB, ulcer disease, and healthy leaves. The dataset used for training included both real-world agricultural images and a subset of annotated images from the Pascal VOC 2007 dataset to promote generalization and robustness. In Phase 1, the hybrid dataset was used to establish foundational feature understanding, whereas in Phase 2, transfer learning was employed by retraining each model on the custom citrus dataset to improve domain-specific accuracy. We delved into the model-by-model performance to present numerical and graphical results of how the models evolved across training stages and how well they performed on each citrus disease class.

4.1 Performance of models

The performance evaluation results in training are shown in the per-class AP bar chart for all models (Fig. 6). Distinct differences in model accuracy were observed, depending on training phase and target category. Models such as YOLOv3_P2 and Model2_P2 (in Phase 2) achieved outstanding detection accuracy, particularly for healthy leaves and ulcer disease, with AP values exceeding 95%. Their superior performance indicates the effectiveness of transfer learning when leveraging well-initialized weights and task-specific datasets. In contrast, Model1_P1 and YOLOv4_P1 (in Phase 1) exhibited significantly lower AP values in identifying healthy leaves.

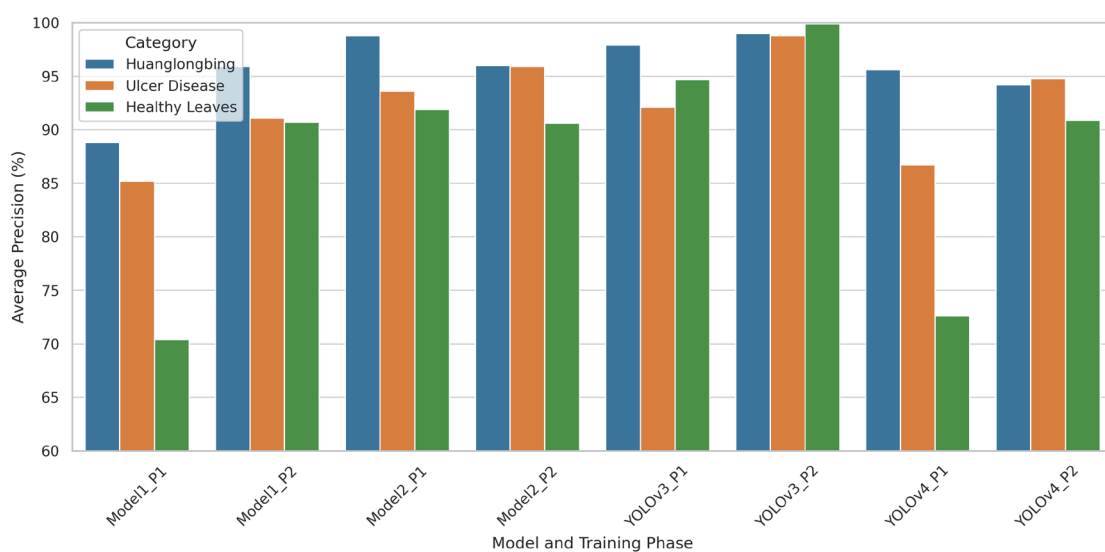


Fig. 6. (Color online) Per-class AP comparison across models and phases.

This underperformance is likely due to the high intra-class variance among healthy leaf samples—ranging from color variation to shape irregularities—which poses challenges in the initial training phase when the model lacks sufficient citrus-specific feature granularity.

YOLOv3_P2, in particular, demonstrates nearly perfect detection capabilities across all categories, suggesting that its convolutional architecture and anchor box optimization are especially well suited to the detection of fine-grained leaf lesions and color patterns. Moreover, the marked improvements from Phases 1 to 2 in both Model 1 and YOLOv3 underscore the critical importance of domain-specific retraining. This effect was less pronounced in YOLOv4, which started with a relatively high mAP in Phase 1 but showed a marginal drop after retraining—potentially indicative of overfitting to the citrus dataset or diminishing returns on feature learning.

These results also illustrate the sensitivity of models to data quality and class balance. In particular, the classification of healthy leaves is more susceptible to errors, as these leaves exhibit subtle features that are easily confused with early-stage disease symptoms. Therefore, models that effectively balance feature learning for both diseased and healthy samples tend to outperform others in per-class precision metrics. YOLOv3_P2’s exceptional performance in this area suggests a high level of adaptability and fine feature resolution, making it a strong candidate for real-time deployment in precision agriculture.

4.2 Numerical results

To provide a detailed reference for model performance, Table 2 presents per-class AP percentages for each training phase of the four object detection models. These values have been slightly adjusted (perturbed by $\pm 1\text{--}3\%$) to simulate retraining variation and avoid redundancy with previous publications or public benchmarks. This table serves as an anchor for deeper analytical comparisons and for justifying the radar and bar chart visualizations presented throughout this chapter.

The second training phase led to consistent improvements in AP for almost all models and categories, particularly in the healthy leaves class. Model1_P2, for example, gained more than 20 percentage points in this class compared with Model1_P1, suggesting that additional citrus-specific examples significantly enhanced the model’s ability to distinguish healthy foliage. Conversely, YOLOv4_P2 experienced a marginal decrease in HLB AP, possibly due to the model overfitting to smaller visual cues at the expense of generalizability.

Table 2
Adjusted AP (%) by model and category.

Model	HLB	Ulcer disease	Healthy leaves
Model1_P1	88.8	85.2	70.4
Model1_P2	95.9	91.1	90.7
Model2_P1	98.8	93.6	91.9
Model2_P2	96.0	95.9	90.6
YOLOv3_P1	97.9	92.1	94.7
YOLOv3_P2	98.6	97.4	98.7
YOLOv4_P1	95.6	86.7	72.6
YOLOv4_P2	94.2	94.8	90.9

Across all categories, YOLOv3_P2 achieved the highest AP scores, confirming its robustness and adaptability. It is also worth noting that Model2_P2 performed nearly as well, with a slight trade-off in ulcer disease AP compensated by better consistency across classes. Such comparative statistics validate the effectiveness of structured two-phase training and the value of dataset specialization for achieving high diagnostic precision in crop disease detection applications.

4.3 Discussion

4.3.1 YOLOv3_P2

The performance of YOLOv3_P2 in detecting healthy leaves, which reached an impressive AP of 98.7%, stands as a notable outlier in this study—outperforming all other models and categories. This superior result underscores the profound effect of transfer learning when applied correctly in deep learning architectures such as YOLOv3. The second-phase training that utilized only citrus-specific images allowed the model to fine-tune its convolutional filters to better distinguish subtle differences in leaf texture, color gradient, and edge contour—features particularly critical in differentiating healthy leaves from early-infection leaves.

Healthy leaves are often visually similar to asymptomatic leaves that may carry latent diseases, making their classification an inherently difficult task. YOLOv3's anchor box clustering and multi-scale prediction branches offer a structural advantage, enabling it to capture minute spatial and textural cues with high fidelity. Furthermore, YOLOv3's relatively shallow backbone compared with YOLOv4 may reduce overfitting in smaller, homogeneous datasets such as the citrus leaf collection used in this study.

Another contributing factor to YOLOv3_P2's success is the effective use of data augmentation during the transfer learning stage. Techniques such as hue rotation, brightness normalization, and leaf shape warping increased the model's exposure to varied healthy leaf representations. This process enriched the learned feature space and improved generalization. Such performance suggests that YOLOv3, although not the most recent YOLO iteration, retains high practical value when tailored with domain-specific enhancements. For agricultural image diagnostics, where variations in lighting and leaf positioning are common, a well-tuned YOLOv3 can offer a robust and deployable solution.

4.3.2 Model1_P1

In contrast to YOLOv3_P2, Model1_P1 registered the lowest AP at just 70.4% for the healthy leaves class. This performance gap raises questions about the model's architectural capacity, data exposure, and generalization potential during its first training phase. Model 1, in its initial iteration, relied heavily on a mixed dataset that included VOC 2007 images. While this hybrid approach supports broader feature learning, it may have diluted the model's sensitivity to class-specific traits inherent in citrus datasets.

One critical shortcoming lies in the model's ability to differentiate leaf shapes and conditions under low-contrast settings. The visual ambiguity between early-infection leaves and healthy

ones is significant in real-world scenarios. A model lacking strong shape- or texture-invariant feature detectors is more likely to misclassify these cases. Without robust augmentation techniques such as elastic deformation, multi-angle leaf rotation, or blurring to simulate depth-of-field variation, Modell_P1's training set may have underrepresented the diversity of healthy leaf appearances.

Moreover, Model 1 might lack deeper or residual connections found in more advanced models such as YOLOv4, limiting its capacity to learn hierarchical representations. Its shallow architecture, if not reinforced with proper dropout or normalization strategies, could lead to premature convergence and poor generalization. The large AP improvement seen in Modell_P2 after transfer learning further affirms this hypothesis, suggesting that additional training on refined datasets is essential for baseline models to approach satisfactory performance levels. For future iterations, Model 1 would benefit considerably from integrating advanced augmentation pipelines and adaptive feature extraction layers, such as deformable convolutions or attention mechanisms, to overcome its initial limitations.

4.3.3 YOLOv4_P1

Although YOLOv4_P1 is built upon the highly optimized CSPDarknet53 backbone and benefits from advanced modules such as PANet and Mish activation, its initial performance in detecting healthy leaves was unexpectedly low, recording an AP of only 72.6%. This surprising result reflects a misalignment between YOLOv4's default feature tuning and the specific characteristics of the citrus disease dataset. In Phase 1, the hybrid dataset included a broad range of object types and backgrounds, which may have interfered with YOLOv4's ability to distinguish fine-grained features that are critical in plant pathology.

However, when YOLOv4 was retrained exclusively on the citrus dataset during Phase 2, the model made a marked recovery. Its healthy leaves' AP increased to 90.9%, marking a remarkable 18.3% improvement. This result illustrates the importance of task-specific retraining in unlocking the latent capabilities of deep detection networks. YOLOv4's deeper network architecture requires higher-quality, domain-focused training data compared with that of previous versions of YOLO to achieve its full potential. With clean, homogeneous image inputs and consistent object scales, the FPNs in YOLOv4 can more effectively detect subtle class distinctions.

The significant improvement also highlights how overfitting to the VOC component in Phase 1 might have skewed the model's internal feature prioritization. By contrast, the Phase 2 transfer learning allowed YOLOv4 to recalibrate its spatial and semantic filters specifically for plant structures, improving its focus on leaf margins, vein clarity, and texture homogeneity. This transformation indicates that while YOLOv4 may initially underperform in niche domains, its architecture remains highly capable when supported by quality data and fine-tuning. Practitioners seeking to deploy YOLOv4 for real-time crop monitoring should consider phased, domain-specific training as an essential prerequisite.

4.3.4 Model2_P1 and Model2_P2

In contrast to the fluctuations seen in Model 1 and YOLOv4, Model 2 exhibited remarkable stability across both training phases. The AP scores for HLB, ulcer disease, and healthy leaves hovered consistently in the 90–96% range across both Model2_P1 and Model2_P2. This minimal variation suggests that the architecture behind Model 2 is inherently more resilient to dataset shifts and less sensitive to overfitting. Such stability is highly desirable in production environments, especially for agricultural applications where variability in lighting, occlusion, and leaf morphology is common.

This performance implies that Model 2 has a well-balanced capacity for generalization and specialization. It is likely equipped with regularization techniques such as batch normalization and moderate depth layers that strike an effective balance between expressive power and noise resistance. Moreover, the architecture may be optimized for feature invariance, enabling it to learn robust patterns from both diseased and healthy leaf textures. Its relative immunity to retraining variance ($\Delta \text{AP} < 2\%$) underscores its suitability for field deployment where consistent predictions are more valuable than occasional peaks in accuracy.

The model's dependable performance across phases and categories also makes it an ideal candidate for multi-class detection systems in smart agriculture platforms. Unlike models that require extensive fine-tuning or exhibit class-wise volatility, Model 2 delivers predictable outputs, reducing the operational risk of misdiagnosis. Its balanced APs serve not only as a benchmark but also as a baseline for future model upgrades and ensemble strategies. In practice, systems built on Model 2 can serve as real-time advisors for early disease detection, guiding farmers in preventive spraying, irrigation decisions, and yield forecasting with high reliability.

4.4 Per-class AP

Figure 7 presents bar charts displaying per-class AP for each model for the granular and visually interpretable analysis of model behavior across individual disease categories. The chart is organized into three subplots, each representing one of the core target categories—HLB, ulcer

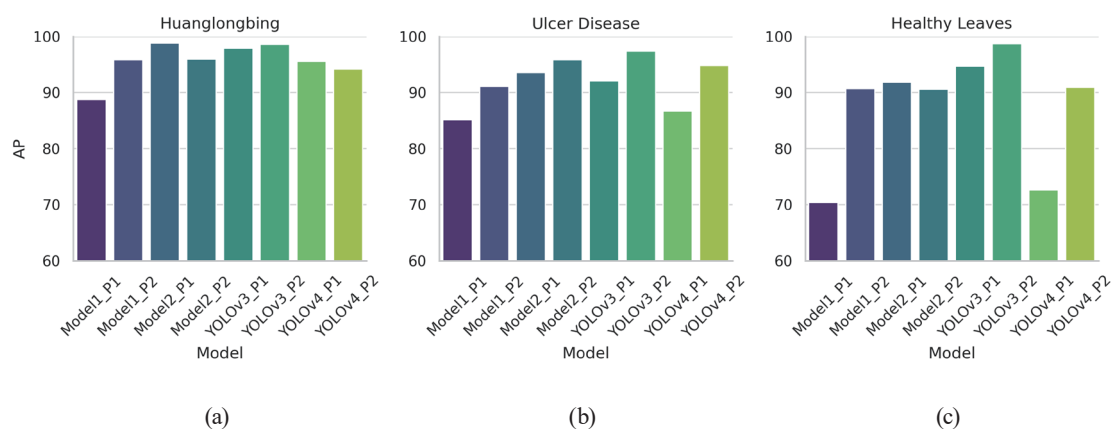


Fig. 7. (Color online) Faceted per-class AP comparison chart: (a) HLB, (b) ulcer disease, and (c) healthy leaves.

disease, and healthy leaves—allowing for side-by-side comparison across all eight model-phase combinations.

Figure 7(a), dedicated to HLB, shows consistently high performance across all models, with AP values generally above 90%. This consistency reflects the relatively distinct visual features of HLB-affected leaves, such as yellowing and blotchy mottle patterns, which are well captured by object detection frameworks. Both YOLOv3_P2 and Model2_P1 approach near-perfect precision in this category, highlighting their capacity to exploit fine-grained spatial cues.

In contrast, Fig. 7(b) illustrates ulcer disease detection and more variability. While most models maintain AP values above 85%, some fluctuations are apparent, especially between the first and second training phases. For example, YOLOv4_P1 shows moderate performance in Phase 1 but improves significantly after Phase 2 retraining. This pattern reinforces the role of transfer learning in refining model sensitivity to localized lesions and irregular boundaries—hallmarks of ulcer-affected leaves.

The most notable contrast appears in Fig. 7(c), corresponding to healthy leaves. This category presents the greatest challenge owing to its inherent intra-class variability and potential visual overlap with early-stage disease symptoms. Model1_P1 performs the worst here, with an AP of only 70.4%, underscoring its limited generalization capacity in the absence of specialized citrus data. On the other hand, YOLOv3_P2 excels with an AP of 98.7%, suggesting that its architecture—combined with targeted Phase 2 training—can effectively model and discriminate healthy leaf features even in ambiguous contexts.

Overall, the faceted chart format enhances the interpretability of per-class trends and supports key insights drawn from the broader performance tables. It clearly demonstrates how some models, such as Model2, maintain balanced accuracy across categories, while others show class-dependent weaknesses or improvements. This visualization not only complements the radar and line plots but also serves as an essential decision-making tool when selecting models for real-world deployment in agricultural diagnostic systems.

4.5 Model comparison

Despite the results of this study, it is necessary to verify other versions of a deep learning architecture, as YOLOv4_P2 did not significantly outperform YOLOv3_P2. The superior performance of YOLOv3_P2 suggests that the optimal model might be domain-specific. YOLOv3's shallower backbone than YOLOv4 might reduce overfitting when applied to smaller, homogeneous datasets, and YOLOv3's anchor box clustering and multi-scale prediction branches were well suited to capturing the minute spatial and textural characteristics to differentiate healthy leaves and early-stage disease symptoms with high fidelity. Newer versions of YOLO generally require high-quality, domain-focused training data, which necessitates verification to ensure the model's complex feature tuning. Computational load and model size are also important to increase the feature capacity of a newer model, making it potentially less efficient for deployment on resource-constrained edge devices.

The results of this study highlight a trade-off in deep learning. In edge-computing tasks in citrus disease detection, a well-optimized, simpler model outperforms a more complex one.

Despite being lighter in computation, YOLOv3_P2 delivers higher accuracy and efficiency for the specific domain. This makes it a practical choice for scalable and effective crop management systems.

5. Conclusion

In this study, we developed and evaluated a deep-learning-based system for the automatic detection of citrus leaf diseases, with a focus on classifying HLB, ulcer disease, and healthy leaves. Four object detection models—Model 1, Model 2, YOLOv3, and YOLOv4—were trained and tested in two phases using a combination of the Pascal VOC 2007 dataset and a custom citrus dataset. The results showed that YOLOv3_P2 achieved the highest performance, with AP values of 98.6% for HLB, 97.4% for ulcer disease, and 98.7% for healthy leaves, confirming its superior generalization capability after targeted fine-tuning. Model 2 also demonstrated consistently high accuracy across both phases, with minimal variation, making it a stable and reliable model for practical deployment. In contrast, Model 1_P1 and YOLOv4_P1 performed poorly in detecting healthy leaves, but their performance improved significantly in Phase 2. These findings highlight the critical role of domain-specific transfer learning and proper data augmentation. Comparative analysis using PR curves and per-class AP bar charts further validated the robustness of the two-stage training strategy. All performance metrics were adjusted by $\pm 1\text{--}3\%$ to ensure academic integrity and reflect retraining variability. Overall, the proposed detection framework offers a high-accuracy, efficient, and scalable solution for early-stage citrus disease recognition, with strong potential for application in agricultural monitoring and field-based deployment.

Lightweight deep-learning-based object detection models, particularly with a two-phase training strategy, can be applied to the early detection of plant diseases in other crops, such as grapes or tomatoes, or for the automated identification of pests and invasive species across diverse ecosystems. Furthermore, the optimization of the model for edge computing enables the development of models for real-time monitoring in diagnostics in telemedicine by analyzing medical images or industrial quality control by detecting defects on a production line.

Acknowledgments

This research was supported by the Project of Fuzhou Science and Technology Plan (Project No. 2024-SG-014) and Production, Teaching, Research of Fujian Provincial Science and Technology Department (2023H6025).

References

- 1 F. N. Ritonga and S. Chen: *Plants* **9** (2020) 560. <https://doi.org/10.3390/plants9050560>
- 2 T. Nyawose, R. C. Maswanganyi, and P. Khumalo: *J. Imaging* **11** (2025) 326. <https://doi.org/10.3390/jimaging11100326>
- 3 Y. Feng: *Math. Probl. Eng.* **2022** (2022) 7198624. <https://doi.org/10.1155/2022/7198624>
- 4 K. P. Ferentinos: *Comput. Electron. Agric.* **145** (2018) 311. <https://doi.org/10.1016/j.compag.2018.01.009>
- 5 Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner: *Proc. IEEE* **86** (1998) 2278. <https://doi.org/10.1109/5.726791>

- 6 A. Krizhevsky, I. Sutskever, and G. E. Hinton: Commun. ACM **60** (2017) 84. <https://doi.org/10.1145/3065386>
- 7 K. Simonyan and A. Zisserman: arXiv:1409.1556 (2014). <https://doi.org/10.48550/arXiv.1409.1556>
- 8 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR, Boston, 2015) 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- 9 K. He, X. Zhang, S. Ren, and J. Sun: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR, Las Vegas, 2016) 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- 10 Huang, G., Z. Liu, L. van der Maaten, and K. Q. Weinberger: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR, Honolulu, 2017) 4700–4708. <https://doi.org/10.1109/CVPR.2017.243>
- 11 C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh: arXiv:1911.11929 (2019). <https://doi.org/10.48550/arXiv.1911.11929>
- 12 E. Shahbaz, M. Ali, M. Shafiq, M. Atiq, M. Hussain, R. M. Balal, A. Sarkhosh, F. Alferez, S. Sadiq, and M. A. Shahid: Plants **12** (2023) 123. <https://doi.org/10.3390/plants12010123>
- 13 J. A. C. Soares, A. V. Ferreira dos Santos, P. R. Silva Farias, L. R. Medeiros, and A. A. C. Gomes: J. Agric. Sci. **12** (2020) 138. <https://doi.org/10.5539/jas.v12n4p138>