Computer Vision-Based Methods for Plant Disease Recognition

Zhong Kairui

Hainan Vocational University of Economics and Business, Haikou 571127, Hainan, China

Abstract: To address the low efficiency and high subjectivity of manual diagnosis, this paper analyzes the task framework, data foundations, and principal computer vision methods used for plant disease recognition. The findings indicate that deep learning has substantially improved classification accuracy for images acquired under controlled conditions. However, occlusion, co-occurring diseases, class imbalance, and domain shift in natural environments continue to constrain practical application. Future research should strengthen cross-regional dataset development; integrate object detection, image segmentation, self-supervised learning, and multimodal information; and improve system reliability through model calibration, uncertainty estimation, and external field testing.

Keywords: computer vision; plant disease; deep learning; image recognition; smart agriculture

Introduction

Plant disease symptoms are often subtle during the early stages of infection. Different diseases may exhibit similar symptoms, while the manifestations of the same disease may vary among cultivars and growth stages. Conventional diagnosis relies on field observations by plant-protection specialists and is therefore limited in efficiency and spatial coverage. Although molecular testing offers relatively high specificity, it requires sample collection, reagents, and appropriate laboratory facilities. Computer vision can extract information on color, texture, shape, and spatial distribution from images of leaves, stems, fruits, or crop canopies, providing a technical foundation for low-cost, noncontact, and large-scale monitoring.

In recent years, the research paradigm has shifted from manually engineered features to end-to-end learning with convolutional neural networks (CNNs) and has subsequently expanded to object detection, semantic segmentation, vision Transformers, and lightweight deployment. Nevertheless, the extremely high accuracy achieved in some studies using closed-category datasets with uniform backgrounds and centrally positioned leaves does not necessarily indicate an ability to recognize early-stage lesions reliably under field conditions. Drawing on six traceable studies, this paper discusses task definitions, model development, evaluation methods, and application constraints and proposes a pathway from “high dataset accuracy” to “reliable field diagnosis.”

1 Tasks and Data Foundations of Visual Plant Disease Recognition

1.1 Task Levels and Technical Boundaries

Visual plant disease recognition comprises at least four levels of tasks. The first is image classification, which determines the crop–disease category to which an entire image belongs. It is suitable for scenarios involving a single leaf, a single disease, and a clearly defined subject. The second is object detection, which simultaneously outputs the classes and bounding boxes of lesions or diseased leaves in complex backgrounds and is therefore suitable for images containing multiple objects. The third is semantic or instance segmentation, which localizes lesions at the pixel level and can subsequently be used to calculate lesion-area proportions and disease severity. The fourth is temporal monitoring, which uses sequential images to assess the rate of disease progression and the effectiveness of interventions.

Visual recognition identifies visible symptoms rather than the pathogens themselves. Nutrient deficiencies, phytotoxicity, and drought or waterlogging stress may produce chlorosis or necrosis resembling disease symptoms, while multiple pathogens may infect a plant simultaneously. Model outputs should therefore be positioned as screening results or diagnostic support. Decisions concerning pesticide application should additionally consider agronomic information, meteorological conditions, and laboratory testing when necessary.

1.2 Visual Representation of Disease Symptoms

Conventional methods generally characterize symptoms from three perspectives. Color features represent chlorosis, browning, and fungal growth and are commonly extracted from the RGB, HSV, or Lab color spaces. Texture features describe lesion roughness, directionality, and repetitive structures and can be obtained using gray-level co-occurrence matrices, local binary patterns, and related descriptors. Shape features characterize lesion boundaries, circularity, area, and leaf-margin deformation. Through multilayer convolution or attention mechanisms, deep networks automatically learn hierarchical representations ranging from edges and textures to composite lesion patterns, thereby reducing their dependence on manually selected features.

More complex symptom representations are not necessarily more reliable. If a particular category consistently co-occurs with a specific background, scale marker, or shooting angle in the dataset, the model may learn irrelevant cues. Saliency maps, occlusion tests, or counterfactual samples should therefore be used to examine the regions on which the model focuses and to confirm that its decisions are based on leaves and lesions rather than soil, hands, or labels.

1.3 Data Acquisition and Annotation Systems

Data quality determines the upper limit of model performance. Controlled datasets facilitate the acquisition of clear and uniform images but cannot adequately represent natural illumination, leaf overlap, wind-induced blur, and complex backgrounds. Images collected from the internet offer greater diversity, but their label provenance, duplication, and copyright status must be reviewed. Fixed-site field acquisition most closely represents actual applications but is constrained by seasonality and disease incidence.

PlantDoc contains 2,598 images covering 13 plant species and 17 disease classes. Its annotation required approximately 300 person-hours, illustrating the high cost of constructing natural-scene datasets [5].

Annotations should be reviewed by plant-protection specialists. Records should include the crop cultivar, growth stage, plant organ, location, date, method used to confirm the pathogen, and disease severity. Data should be partitioned at the level of individual plants, plots, or acquisition batches. Rotated or cropped versions of the same leaf must not appear simultaneously in the training and test sets, as this would cause data leakage. For class imbalance, resampling and loss weighting should be supplemented by the continuous collection of samples from minority classes and easily confused categories.

2 Evolution of Plant Disease Recognition Methods

2.1 Conventional Image Processing and Machine-Learning Methods

A conventional recognition pipeline consists of image preprocessing, lesion segmentation, feature extraction, and classification. Preprocessing techniques such as white balancing, filtering, and contrast enhancement are used to reduce differences in image acquisition. Segmentation can be performed using color thresholds, clustering, region growing, or active contour models. Common classifiers include support vector machines, random forests, k-nearest-neighbor algorithms, and shallow neural networks. These methods require relatively little computation, and their features are readily interpretable. They therefore remain valuable for single-crop tasks involving limited samples and stable backgrounds.

Their principal limitation is that errors are propagated from one processing stage to the next. Color thresholds are sensitive to variations in sunlight and cameras, while manually designed texture features cannot readily cover multiple diseases and different stages of disease development. When leaves occlude one another or lesion boundaries are indistinct, the stability of a segmentation-first, classification-second pipeline decreases. Conventional features are therefore more suitable as complements to deep features or for systems with limited computing resources and strictly controlled operating environments.

2.2 Deep-Learning-Based Recognition Methods

CNNs enable end-to-end classification through local connectivity, weight sharing, and multilayer feature learning. When only a small number of samples are available, transfer learning is generally performed using models pretrained on ImageNet, with either the upper layers fine-tuned or the network layers progressively unfrozen.

Mohanty et al. trained models using 54,306 leaf images acquired under controlled conditions and achieved a test accuracy of 99.35% across 38 crop–disease combinations [1]. Ferentinos used 87,848 images representing 58 classes, and the best-performing model achieved an accuracy of 99.53% on 17,548 images not used for training [2]. These findings demonstrate the feasibility of deep visual representations but primarily reflect performance within the corresponding data distributions.

For field images, object detectors are better able than whole-image classifiers to handle multiple leaves and complex backgrounds. Faster R-CNN is representative of two-stage methods, which generally achieve high localization accuracy but impose relatively high inference costs. One-stage methods directly predict classes and locations, making their speed more suitable for mobile devices and robots.

Segmentation networks can output lesion contours for disease-severity quantification, although pixel-level annotation is expensive. Vision Transformers are well suited to modeling long-range dependencies, while hybrid CNN–Transformer architectures combine local texture information with global morphology. With small agricultural datasets, however, transfer learning, regularization, or self-supervised pretraining remains necessary to control overfitting.

3 Model Development, Evaluation, and Empirical Comparison

3.1 Data Preprocessing and Training Strategies

Before modeling, datasets should be examined for duplicate samples, incorrect labels, and class distributions, and an independent test set containing no augmented images should be retained. Geometric augmentation may include flipping, rotation, cropping, and scaling. Color augmentation should remain within a range that does not destroy pathological characteristics, as excessive color shifts may alter critical evidence such as chlorosis and rust-colored lesions. Methods such as CutMix and MixUp can improve generalization, but their synthetic boundaries should not be misinterpreted as authentic lesions.

Cross-entropy loss, focal loss, or class-weighted loss can be used during training to address difficult examples and class imbalance. Model selection should be performed on the validation set, while the test set should be used only for the final evaluation. If cross-regional application is intended, leave-one-location-out or leave-one-season-out external validation should be conducted. If the same plant is photographed repeatedly, the data should be partitioned by plant. Hyperparameters, random seeds, image resolution, and inference hardware should all be reported to ensure reproducibility.

3.2 Evaluation Metrics and Experimental Design

Accuracy is appropriate for balanced, closed-set test datasets, but it may conceal false negatives when identifying minority disease classes. Precision is defined as ($P=TP/(TP+FP)$), recall as ($R=TP/(TP+FN)$), and the F_1-score as ($F_1=2PR/(P+R)$). Object-detection studies should also report mean average precision (mAP) at a specified intersection-over-union threshold, while segmentation studies may report intersection over union (IoU) or the Dice coefficient.

Recall is generally assigned greater importance in practical early-warning systems, whereas automated pesticide application must also control false positives. Decision thresholds should therefore be determined according to the costs associated with the intended application.

Evaluation should not be limited to mean performance. A class-specific confusion matrix, confidence intervals, calibration error, model parameter count, inference time per image, and energy consumption should also be reported. For previously unseen diseases, abiotic stresses, and low-quality images, the model should be permitted to reject the input or refer it to an expert instead of forcibly assigning it to a known class. External-test performance and the reliability of confidence estimates provide more meaningful evidence of practical value than a single internal accuracy score.

3.3 Comparison of Representative Studies

Table 1 summarizes representative evidence obtained under different data conditions. Because the studies differ in species, class definitions, task formats, and data-partitioning methods, their reported accuracies cannot be used directly to rank the models.

Table 1. Data and results from representative visual plant disease studies

Study	Images and acquisition setting	Task scope	Reported results
Mohanty et al. [1]	54,306 images acquired under controlled conditions	14 crop species and 38 crop–disease combinations	Accuracy of 99.35% and a mean F_1-score of 0.9934
Ferentinos [2]	87,848 images, including a test set of 17,548 images	25 plant species and 58 classes	Best recognition rate of 99.53%
Ramcharan et al. [3]	2,756 field images of cassava acquired in Tanzania	Three diseases and two pest or healthy-status categories	Overall independent-test accuracy of 93%
Fuentes et al. [4]	Approximately 5,000 tomato images acquired on farms in South Korea	Detection and localization of nine disease and pest classes	Compared Faster R-CNN, R-FCN, and SSD and achieved multi-object localization
Singh et al. [5]	2,598 internet-sourced natural-scene images	13 plant species and 17 disease classes	Training with the dataset increased classification accuracy by as much as 31%

As shown in Table 1, controlled-image classification has achieved extremely high accuracy, whereas field studies are more capable of revealing interference from backgrounds and variations in symptom appearance. Based on an analysis of a database containing nearly 50,000 images covering 21 plant species and 171 diseases, Barbedo reported that backgrounds, image-acquisition conditions, symptom segmentation, and data diversity all affect recognition performance in real-world environments [6]. Research papers should therefore prioritize explaining the difference between the test dataset and the intended deployment environment rather than using an accuracy exceeding 99% as the sole indication of technological maturity.

4 Key Constraints on Field Application

4.1 Domain Shift and Open-Environment Recognition

Differences between training and application data in cameras, regions, cultivars, seasons, and management practices create domain shift. Relationships between colors and backgrounds learned in one location may become invalid when the model is transferred to another region. During the early stage of a disease, symptomatic pixels occupy only a small proportion of the image and the symptoms are weak. During later stages, the symptoms may converge with those of other diseases. Raindrops, dust, insect damage, and nutrient stress further increase interclass confusion.

Field environments may also contain emerging diseases not represented in the training set. Conventional closed-set classifiers can assign such inputs to incorrect classes with apparently high confidence.

Addressing these problems requires multisite and cross-seasonal external testing. Color constancy, domain adaptation, and style perturbation can be used to reduce image-acquisition differences. Contrastive learning or self-supervised learning can exploit large volumes of unlabeled images. Open-set recognition and uncertainty estimation should be introduced to enable the system to identify “unknown” cases. Model updates should also prevent catastrophic forgetting and document the model version, training data, and applicable geographical areas.

4.2 Lightweight Deployment, Trustworthy Interpretability, and Decision Risks

Mobile devices require a balance among accuracy, speed, memory consumption, and energy use. Models can be compressed through depthwise separable convolution, structured pruning, knowledge distillation, and low-bit quantization. However, recall for minority classes must be reevaluated after compression.

A hierarchical architecture in which initial screening is performed at the edge and complex reviews are completed in the cloud can balance response speed with model capability. In areas with unstable network connectivity, offline inference and delayed synchronization should remain available.

Interpretability involves more than displaying heat maps. It is necessary to verify that heat maps consistently cover lesions and to test whether predictions decrease appropriately after lesions are occluded. System outputs should include candidate diseases, confidence estimates, key image regions, and applicability boundaries rather than directly generating definitive pesticide-use instructions. Misdiagnosis may delay disease management or cause pesticide misuse. Expert–model collaborative trials, failure-mode reviews, and data-privacy assessments should therefore be conducted before deployment.

5 Methodological Optimization and Development Trends

For practically usable systems, a closed-loop framework of “acquisition–diagnosis–feedback–retraining” should be established. At the data level, RGB images should be collected using consistent protocols. For major crops, RGB imaging should be integrated with near-infrared and thermal-infrared data as well as meteorological, soil, and cultivation records. Separate labels should be assigned to mixed-pathogen infections and abiotic stresses.

At the model level, multitask architectures combining diseased-leaf detection, lesion segmentation, and disease classification should be adopted so that spatial localization, severity estimation, and class prediction can constrain and reinforce one another. Self-supervised pretraining and few-shot learning should be used to reduce dependence on manual annotation.

At the evaluation level, a hierarchical benchmark comprising controlled datasets, natural-scene datasets, cross-regional datasets, and open-set datasets should be established. Macro-averaged F_1-scores, class-specific recall, mAP, calibration error, and on-device efficiency should be reported.

At the application level, model confidence can be combined with epidemiological conditions favorable to particular diseases. For common diseases identified with high confidence, the system may provide management recommendations. Cases involving low confidence or high-risk diseases should be referred to experts and confirmed through laboratory testing to complete the diagnostic loop.

Future research should also promote privacy-preserving cross-regional collaborative training, continual learning, and causal validation, enabling recognition systems to evolve from “image-similarity matching” into trustworthy decision-support tools that incorporate plant-pathology knowledge.

Conclusion

Computer vision has substantially increased the level of automation in plant disease screening, and deep learning has demonstrated strong feature-learning capabilities on standardized datasets. However, high accuracy under controlled conditions cannot replace external field validation. Data diversity, unknown diseases in open-set environments, combined stresses, and device constraints remain major barriers to large-scale application.

Future research should be founded on high-quality cross-regional data and should integrate classification, object detection, segmentation, and multimodal information. Generalization, calibration, interpretability, and on-device efficiency should also be incorporated into a unified evaluation framework. Only by establishing a diagnostic loop that incorporates expert participation and manageable uncertainty can visual recognition genuinely support precision plant protection.

References

[1] Mohanty S P, Hughes D P, Salathé M. Using deep learning for image-based plant disease detection [J]. Frontiers in Plant Science, 2016, 7: 1419.

[2] Ferentinos K P. Deep learning models for plant disease detection and diagnosis [J]. Computers and Electronics in Agriculture, 2018, 145: 311–318.

[3] Ramcharan A, Baranowski K, McCloskey P, et al. Deep learning for image-based cassava disease detection [J]. Frontiers in Plant Science, 2017, 8: 1852.

[4] Fuentes A, Yoon S, Kim S C, et al. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition [J]. Sensors, 2017, 17(9): 2022.

[5] Singh D, Jain N, Jain P, et al. PlantDoc: A dataset for visual plant disease detection [C]// Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. New York: ACM, 2020: 249–253.

[6] Barbedo J G A. Factors influencing the use of deep learning for plant disease recognition [J]. Biosystems Engineering, 2018, 172: 84–91.