Journal of Agricultural Engineering

https://www.agroengineering.org/

Recent advances in pest and disease recognition: a comprehensive review

Honglin Liu,[1] Bisheng Zhan,[1] Ruitong Fang,[1] Yi Zhang,[1] Yujiao Ma,[1] Ze Shen,[1] Qirong Mao[1,2,3]

[1]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang
[2]Jiangsu Engineering Research Center of Big Data Ubiquitous Perception and Intelligent Agricultural Applications, Zhenjiang
[3]Provincial Key Laboratory of Computational Intelligence and New Technologies in Low-Altitude Digital Agriculture, Zhenjiang, Jiangsu China

Corresponding author: Qirong Mao, School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China. E-mail: mao_qr@ujs.edu.cn

Please cite this article as doi: 10.4081/jae.2025.1776

Submitted: 16 March 2025
Accepted: 4 July 2025

Recent advances in pest and disease recognition: a comprehensive review

Honglin Liu,[1] Bisheng Zhan,[1] Ruitong Fang,[1] Yi Zhang,[1] Yujiao Ma,[1]
Ze Shen,[1] Qirong Mao[1,2,3]

[1]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang
[2]Jiangsu Engineering Research Center of Big Data Ubiquitous Perception and Intelligent Agricultural Applications, Zhenjiang
[3]Provincial Key Laboratory of Computational Intelligence and New Technologies in Low-Altitude Digital Agriculture, Zhenjiang, Jiangsu China

Corresponding author: Qirong Mao, School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China. E-mail: mao_qr@ujs.edu.cn

Abstract
Agricultural pests and diseases pose a severe threat to global food production, making timely and accurate recognition crucial for ensuring crop health and enhancing yields. With the rapid advancement and application of artificial intelligence (AI) across various scientific domains, its potential in pest and disease recognition remains only partially explored. Therefore, we conduct a comprehensive review, focusing on the latest progress in applying machine learning (ML), deep learning (DL), and multimodal technologies to pest and disease recognition in agriculture. It covers state-of-the-art techniques, benchmark datasets, and evaluation metrics relevant to this field. Additionally, the review offers an in-depth understanding of the strengths, challenges, and limitations of these methods. We also highlight several representative studies and conduct a comparative analysis of their performance. Finally, the paper provides detailed insights, proposes potential research directions, and concludes with reflections on future advancements.

Keywords: Crop health; deep learning; machine learning; multimodal technologies; pest and disease recognition.

Introduction

Agricultural pests and diseases significantly threaten global food security, causing severe annual crop losses and disrupting food supply chains worldwide. These losses, which can reach up to 40% of global yields, endanger economic stability and hinder efforts to ensure sustainable agricultural practices in the face of a growing global population (Savary *et al.*, 2019; Junaid and Gokce, 2024; He *et al.*, 2023). Timely and accurately recognizing pests and diseases is critical for mitigating their impact, enhancing crop health, and increasing agricultural productivity (Zhang *et al.*, 2017; Yang *et al.*, 2020; Zhao *et al.*, 2021). Traditional methods, such as manual inspection and expert analysis, have long been used to identify and manage agricultural pests and diseases. While effective in certain localized scenarios, these methods are time-intensive, laborious, and susceptible to human error (Pujari *et al.*, 2015; Lu *et al.*, 2021). Furthermore, their scalability is limited in modern agricultural settings where large-scale monitoring and real-time decision-making are increasingly essential. These limitations underscore the urgent need for automated, efficient, and scalable solutions. With the development of artificial intelligence (AI) technology over the past decade, as shown in Figure 1, the number of related publications and citations has grown rapidly, providing a solid foundation for breakthroughs in pest and disease recognition. Machine learning (ML) techniques, such as support vector machines (SVMs) and random forests (RFs), have been successfully employed to classify diseases using handcrafted features such as color, texture, and shape (Mohanty *et al.*, 2016; Ferentinos, 2018; El-Mesery *et al.*, 2024). Meanwhile, deep learning (DL), particularly convolutional neural networks (CNNs), has revolutionized the field by automatically learning hierarchical features from raw data, achieving state-of-the-art performance in tasks involving large and complex datasets (Liu *et al.*, 2021; Peng *et al.*, 2021; Yang *et al.*, 2022). Advanced architectures, such as ResNet, DenseNet, and EfficientNet, have further enhanced the accuracy and scalability of detection systems (Deng *et al.*, 2022; Jouini *et al.*, 2024; Zhu *et al.*, 2024).

Despite the success of ML and DL approaches in pest and disease detection, most models rely solely on RGB images, which are vulnerable to lighting variations, background interference, and ambiguous visual symptoms. To overcome these challenges, recent studies have introduced multimodal data sources as complementary inputs (De Silva and Brown, 2022; Navaneethan *et* *al.*, 2023; De Silva and Brown, 2023). For example, infrared images can highlight temperature variations associated with disease stress, while environmental factors like humidity and temperature offer contextual insights. This complementary information helps overcome the shortcomings of single-modality systems, enhancing detection accuracy and robustness even in complex and variable conditions (Zhu *et al.*, 2022a; Zhang L *et al.*, 2024; Wei *et al.*, 2023). With the rapid development of computer vision, pest and disease recognition technology has been widely studied and applied in the field of agriculture. Through pest and disease detection, agricultural productivity and efficiency have been significantly improved. However, the current application scope of this technology in agriculture remains limited, and the related algorithms are not yet fully mature. This paper aims to summarize the advancements in pest and disease recognition technology over the past decade, providing valuable insights and references for future research and practical applications. In addition, we provide an overview of pest and disease recognition methods used in agriculture, along with an analysis of the strengths and limitations of each approach. The detailed steps for pest and disease recognition are illustrated in Figure 2. We also explore the current challenges and potential future research directions in this field. This study aims to assist researchers in understanding the current application status of pest and disease recognition algorithms in agriculture, offering valuable insights for further advancements that could drive significant breakthroughs in smart and precision farming. Compared with earlier reviews such as Mohanty *et al.* (2016) and Ferentinos (2018), which primarily focused on the application of CNN-based methods and early-stage deep learning models for plant disease classification, this review provides a more comprehensive and up-to-date survey by incorporating recent advancements in Transformer architectures, multimodal fusion technologies, and real-time edge deployment strategies. Additionally, unlike prior studies that mainly emphasized image-based disease recognition, our review systematically categorizes methods into machine learning, deep learning, and multimodal approaches, and further discusses their evaluation metrics, benchmark datasets, and emerging challenges. By integrating studies up to 2025, this paper offers deeper insights into practical deployment issues, data heterogeneity, and the role of large language models (LLMs) in intelligent agriculture, which are rarely discussed in earlier works. The primary contributions can be summarized as follows:

This paper reviews advancements in pest and disease recognition, focusing on ML, DL, and multimodal techniques. It provides a comprehensive understanding of the state-of-the-art techniques and methodologies used in pest and disease recognition.
Various datasets and performance evaluation metrics related to pest and disease recognition are discussed in detail, providing a comprehensive overview of the resources and criteria utilized in this field.
Highlighting existing research gaps and challenges, and offering a forward-looking perspective on future developments in intelligent agriculture.

Materials and Methods

Pest and disease image preprocessing

Image augmentation, dimensionality reduction, and quality enhancement methods have been effectively applied in crop pest and disease image preprocessing to improve recognition accuracy and robustness (Pei *et al.*, 2022; Zhu *et al.*, 2023). By augmenting and downscaling image data, researchers can optimize model performance and address issues like insufficient training data and overfitting (Zhang Z *et al.*, 2024).

Image augmentation

Common image data augmentation methods include rotation, resizing, cropping, affine transformations, and other operations. By applying these transformations, new training samples can be generated, expanding the dataset and improving the model's adaptability to diverse scenarios. In crop pest and disease recognition, the growth angle of plant leaves and the distribution of pests and diseases are influenced by lighting conditions and viewpoints. Randomly rotating and flipping images allow the model to learn pest and disease features from various angles and orientations, thereby enhancing its robustness and recognition accuracy. For instance, Ramcharan *et al.* (2017) demonstrated the effectiveness of rotation and flipping in enhancing the performance of deep learning models for cassava disease detection, showing that augmented data improves robustness to variations in lighting and orientation. Similarly, Mohanty *et al.* (2016) explored the use of scaling and color transformations for plant disease classification, highlighting that these techniques significantly improve model accuracy, especially when training data is scarce. Further studies, such as the work by Pawara *et al.* (2017), have compared multiple augmentation strategies, including rotation, flipping, and noise addition, and found that combining these methods leads to even greater improvements in model performance. These findings underscore the importance of data augmentation as a critical tool for developing reliable and generalizable pest and disease recognition systems in agriculture.

Image dimensionality reduction

In image recognition tasks, images often contain high-dimensional features, and dimensionality reduction methods enhance model efficiency and accuracy by compressing the feature space (Yao *et al.*, 2021; Khulal *et al.*, 2016; Tahir *et al.*, 2016). Commonly used data dimensionality reduction methods are principal component analysis (PCA), linear discriminant analysis (LDA), and t-distributed stochastic neighborhood embedding (t-SNE). Ali *et al.* (2022) presented a novel approach for crop disease identification using feature fusion and PCA-LDA classification, achieving high accuracy in potato crop leaf disease identification. Shi *et al.* (2017) proposed a spectral vegetation indices-based kernel discriminant approach (SVIKDA) for detecting and classifying pests and diseases in winter wheat, achieving high classification accuracy at both leaf and canopy levels. SVIKDA outperforms traditional methods by effectively addressing redundant information in hyperspectral data, demonstrating reliable performance and transferability in pest and disease detection for precision agriculture.

Image quality enhancement

Pest and disease images often suffer from low quality due to factors like poor lighting, low resolution, and noise, making it difficult to distinguish between subtle differences in crop health. Image quality enhancement is essential to improve clarity and highlight critical features, enabling more accurate detection and analysis of pests and diseases. Li *et al.* (2014) focused on identifying weak signal molecules in plant disease and pest images, using lifting wavelet transform and image recognition techniques to analyze corn pest images. Simulation results demonstrated that the method achieved a reliability of

71.65% for plant disease identification and 76.21% accuracy for edge detection, providing a fast, hardware-friendly solution for plant disease image analysis. Nti *et al.* (2017) presented an automatic plant disease detection system using computer vision techniques, including Gaussian smoothing for noise reduction, to identify affected spots on plant leaves, achieving an overall accuracy of 90.96% based on experimental results.

Pest and disease recognition algorithms

Machine learning methods

Feature extraction is crucial for identifying crop pests and diseases, and during the machine learning stage, manual feature extraction is commonly used. Manual features typically involve low-level information such as shape, color, and texture, with common descriptors including color histograms, color moments, grayscale covariance matrices, and directional gradient histograms (Huang *et al.*, 2018; Yang *et al.*, 2019a). After extracting these features, classifiers such as support vector machines (SVM), RF, and k-nearest neighbor (KNN) clustering are trained and used to identify and classify plant pests and diseases. To illustrate, Yang *et al.* (2019a) found that the synergistic judgment of texture and shape features combined with the decision tree-confusion matrix method can quickly and accurately detect rice diseases, which is crucial for formulating early prevention strategies and effectively controlling rice diseases. Lu *et al.* (2021) proposed a method for identifying tea white star disease and anthrax based on hyperspectral imaging, which employs the machine learning technique of extreme learning machine (ELM). The study demonstrated that the accuracy of disease identification was significantly enhanced when the diseased area was segmented with mask technology and combined with the ELM model. Kale and Shitole (2021) employed the RF, SVM, and KNN algorithms to detect pests and diseases. They found that Multivariate Support Vector Machines (MSVM) exhibited high accuracy in disease classification and detection, but the main challenge is the difficulty of feature extraction. Kumar and E (2022) proposed a prototype for detecting rice plant diseases, including bacterial leaf blight, brown spot, and leaf smut, using machine learning and image processing techniques. The prototype achieved an accuracy of 98.8% in detecting and classifying rice leaf diseases by extracting important features through Discrete wavelet transform (DWT) and applying an adaptive boosting support vector machine (AdaBoostSVM) classifier.

Deep learning methods

The application of deep learning in agriculture has become increasingly widespread (Zhou *et al.*, 2020; Liu *et al.*, 2021). These technologies improve the accuracy of pest and disease recognition and help agricultural producers manage crop health more efficiently, reducing the cost and time of manual detection. In recent years, many innovative deep learning methods have been proposed, covering various models such as CNN, transformer, and recurrent neural networks (RNN), which have greatly promoted the development of crop pest and disease recognition technology (Tian *et al.*, 2022; Wang J *et al.*, 2021).Yang *et al.* (2019b) proposed a model for detecting tea leaf diseases using infrared thermal imaging technology. The model extracts feature parameters through image classification, color recognition, and threshold segmentation, and inputs them into the classifier. By leveraging the characteristics of infrared thermal images, the model effectively improves recognition accuracy. Yang *et al.* (2020) built a model based on CNN and trained it using the Kaggle tea illness dataset. By employing lightweight deep neural networks like NASNet, they further enhanced the accuracy, achieving a high precision of 95.90%, showcasing its advantages in precise disease classification. Zhu *et al.* (2022b) embedded the Transformer encoder into CNNs, using the self-attention mechanism to capture long-distance dependencies between features in the image, extract global features, and introduce Centerloss to optimize the loss function, enhancing class separation and reducing intra-class differences, thus improving recognition accuracy. Chithambarathanu and Jeyakumar (2023) used LSTM to analyze environmental data from weather stations to predict pest outbreaks. RNNs perform excellently with sequence data, effectively capturing long-term dependencies in time-series data, and providing more accurate predictions for pest forecasting. Singh *et al.* (2021) proposed a framework based on deep learning to detect infections from coconut tree pests and diseases. By applying image processing and deep learning technologies, the model detects stem bleeding disease, leaf blight, and red palm weevil infections. It uses a segmentation algorithm to process images and employs a custom-designed deep 2D CNN for training and prediction, with Keras pre-trained CNN models using inductive transfer learning methods for image classification.

Multimodal methods

With the rapid development of smart agricultural technologies, the fast and accurate recognition of pests and diseases has become a key issue in improving agricultural productivity and ensuring food security (Liu and Wang, 2024). Traditional methods primarily rely on a single data source, such as images or text. However, these methods face significant limitations in terms of applicability and accuracy, especially in complex environments. In recent years, multi-modal learning, as a technology that integrates various data sources (e.g., images, text, sensor data, *etc*.), has made significant progress (Cheng *et al*., 2024; Liu Z *et al*., 2024). By combining different types of data sources, multimodal fusion not only improves recognition accuracy but also enhances the robustness of the system. This paper aims to review the application of multi-modal fusion technology in pest and disease identification, focusing on methods such as the fusion of image features, image-text fusion, and the fusion of image and environmental data (Yu K *et al*., 2024).

Wang Y *et al*. (2021) tackled weed recognition tasks by combining multiple image features with a back propagation neural network (BPNN), which significantly reduced computational cost and recognition time while improving overall system efficiency. Meanwhile, Xu *et al*. (2023) enhanced the estimation accuracy of leaf nitrogen content in rice and improved crop health monitoring and early warning capabilities by integrating visual and spectral features and applying a minimum redundancy maximum relevance method. Compared to traditional fusion methods that rely on simple concatenation or handcrafted feature selection, recent multimodal frameworks incorporate attention mechanisms to achieve more precise semantic alignment. These mechanisms typically model cross-modal interactions by using visual features as queries and textual or environmental features as keys and values in a transformer-like attention structure. Through this formulation, the model learns to focus selectively on the most semantically relevant information across modalities, enabling fine-grained alignment between visual cues and external descriptions. For example, Liu and Wang (2024) proposed a framework that integrates image and text, using CNN to extract image features and self-attention model to extract semantic features from text, and then fusing both features to distinguish similar pests and diseases. The ITF-WPI model combines image information with environmental data from sensors and employs a pyramid squeezing attention mechanism to enhance multi-scale feature extraction efficiency, further strengthening the model's computational capability and recognition performance (Dai *et al.*, 2023).

Benchmark datasets

Agricultural pest and disease recognition plays a crucial role in ensuring crop health and optimizing farming practices (Zhang *et al.*, 2022). Early and accurate recognition of pests and diseases can significantly improve crop yield and quality, while also reducing the need for harmful pesticides (Wu *et al.*, 2023). High-quality datasets are foundational for developing effective methods for automatic detection and classification, as they provide the necessary data for training, validating, and testing these models. With the advancement of computer vision techniques, the availability of diverse and well-annotated datasets has become essential in fostering progress in this field (Zhu *et al.*, 2022b). In this section, we review several important datasets used in agricultural pest and disease recognition, categorized by their focus on plant diseases or pests. These datasets vary in the types of data they offer, such as RGB images, thermal images, multispectral data, and more, each serving a specific purpose depending on the agricultural context and the type of pest or disease being detected. In Table 1, we have compiled a list of several commonly used benchmark datasets for pest and disease recognition. This table includes datasets such as Plant Village (Hughes and Salathé, 2015), IP102 (Wu *et al.*, 2019), Rice Leaf Disease (Sethy *et al.*, 2020), Vine Disease (Kerkech *et al.*, 2020), AgriVision (Chiu *et al.*, 2020), CPB (Bollis *et al.*, 2020), WeedNet (Sa *et al.*, 2017) , and CCMT (Mensah *et al.*, 2023). Each dataset is described in terms of the type of data it contains, including RGB images, multispectral images, or thermal images, and the source of the data. To provide further insight, we have also randomly selected a few sample images from each dataset, which are displayed in Figure 3. These samples offer a visual representation of the data and help illustrate the types of images used for training models in pest and disease detection.

Despite the availability of several benchmark datasets, practical challenges remain in constructing high-quality datasets for pest and disease recognition.

First, collecting images under diverse and uncontrollable field conditions (e.g., varying lighting, occlusion by foliage, or motion blur) can lead to inconsistent data quality. Second, accurate annotation requires expert agronomic knowledge, which is time-consuming and labor-intensive. Third, the distribution of samples is often imbalanced, as certain pests or diseases occur rarely or seasonally, making it difficult to obtain sufficient training data for these classes. Moreover, inconsistencies in labeling criteria across datasets may further hinder model generalization. These factors underline the importance of developing robust learning methods that can cope with noisy, sparse, or weakly annotated data.

Evaluation metrics

Metrics based on the confusion matrix

Pest and disease detection can be simplified as a binary classification problem, where the classes are "presence of pests/diseases (1)" and "absence of pests/diseases (0)." The confusion matrix is used to describe the distribution of the model's prediction results. Its structure is shown in Table 2.

1) Precision: precision measures the proportion of correctly predicted positive samples (actual pests/diseases) among all samples predicted as positive by the model. Precision focuses on evaluating the accuracy of the model's predictions. A higher precision indicates fewer false positives when identifying pests/diseases. The formula for precision is as follows:

$$P = \frac{TP}{TP + FP} \quad \text{(Eq. 1)}$$

2) Recall: recall measures the proportion of actual positive samples (pests/diseases) that are correctly predicted by the model. The primary focus of recall is to evaluate the model's ability to comprehensively identify pests/diseases. A higher recall indicates fewer missed cases of pests/diseases (false negatives). The formula for recall is given as:

$$R = \frac{TP}{TP + FN}$$ (Eq. 2)

3) Accuracy: Accuracy reflects the proportion of correctly predicted samples to the total number of samples. It provides an overall evaluation of the model's prediction performance for all samples, including both pest/disease and healthy cases. The formula for accuracy is:

$$Accuracy = \frac{TP}{TP + TN + FP + FN}$$ (Eq. 3)

4) F1-score: F1-score is the harmonic mean of precision and recall, providing a balanced metric when there is a need to balance between precision (reducing false positives) and recall (reducing false negatives). F1-Score is especially useful when the model's predictions need to balance these two aspects. The formula for F1-Score is:

$$F1\_Score = 2 \times \frac{P \times R}{P + R}$$ (Eq. 4)

Metrics based on region overlap

In pest and disease detection tasks, the model is required not only to determine whether pests/diseases are present in the sample but also to accurately locate the regions of pests/diseases. The intersection over union (IoU) is an important metric used to measure the overlap between the predicted region and the ground truth region. The formula for IoU is:

$$IoU = \frac{Area_{predict} \cap Area_{groundtruth}}{Area_{predict} \cup Area_{groundtruth}}$$ (Eq. 5)

IoU quantifies the accuracy of the model in locating the pest/disease region. A higher IoU value indicates a greater degree of overlap between the predicted region and the ground truth region. When the IoU is greater than or equal to a predefined threshold (commonly 0.5), the prediction is considered a valid detection.

Comprehensive evaluation metrics

1) ***Average precision (AP):*** AP represents the average value of precision under different recall rates. It is used to evaluate the detection performance of the model for a single pest or disease category and reflects the overall performance of the model in that category. The formula for AP is:

$$AP = \int_0^1 P(R)dR \quad \text{(Eq. 6)}$$

2) Mean average precision (mAP): mAP is the most commonly used evaluation metric in pest and disease detection tasks. It is the mean of the AP values across all pest and disease categories, and it is used to assess the overall performance of the model in multi-category detection tasks. The formula for mAP is:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad \text{(Eq. 7)}$$

Here, *N* represents the total number of pest and disease categories in the test set. Typically, the calculation of mAP is based on a fixed IoU threshold (e.g., 0.5). A prediction is considered correct only if the IoU between the predicted region and the ground truth region is greater than or equal to the threshold. A higher mAP value indicates that the model has a more balanced detection performance across all categories.

Recent methods and developments
Current representative ML methods

In recent years, significant advancements have been made in the field of crop pest and disease recognition, driven by continuous improvements in machine learning technologies. For pest recognition, Deng *et al.* (2018) proposed a bio-inspired pest image detection and recognition model. They integrated a visual saliency model (SUN) with an enhanced HMAX model, using SIFT to extract rotation-invariant features and local configuration pattern (LCP) for texture features. Classification was performed using SVM, achieving a recognition rate of 85.5% under complex environmental conditions. Kasinathan and Uyyala (2021) developed a pest detection system combining RF and KNN classifiers, leveraging features such as color, texture, and shape to enhance classification performance. This method significantly improved accuracy, achieving over 90% on multiple pest image datasets with 10-fold cross-validation. Next year, Gomes and Borges (2022) proposed a few-shot learning approach for insect pest recognition, using the IP-FSL dataset with 97 adult pest classes and 45 early-stage pest classes. The method achieved 86.33% accuracy for adults and 87.91% for early stages, showing promise for early pest detection in crop scenarios. Plant diseases are another significant factor influencing crop yield. Early detection and effective management of these diseases can enhance both the quality and productivity of agricultural outputs. Yang *et al.* (2019a) combined texture and shape features with the decision tree-confusion matrix method, enabling rapid and accurate detection of rice diseases, which is crucial for formulating early prevention strategies and effectively controlling rice diseases. Bainalwar *et al.* (2023) designed an intelligent plant disease detection system utilizing GLCM feature extraction combined with KNN and Random Forest classifiers. The system achieved real-time classification and showed high accuracy in distinguishing healthy and diseased leaves. Kethineni and Pradeepini (2023) introduced a model integrating Genetic Algorithms (GA) with KNN for leaf disease classification, optimizing feature selection to improve performance. To present these representative ML methods more intuitively, we have organized and summarized them, as shown in Table 3.

Current representative DL methods

The application of deep learning in crop pest and disease recognition has rapidly expanded, with its primary advantage being the automatic extraction of features from pest and disease images, facilitating efficient recognition and classification of disease types. This eliminates the need for manual feature engineering, significantly enhancing detection accuracy and efficiency. A region-based CNN model proposed by Rehana *et al.* (2023) optimized the detector network and introduced a lightweight structure to enhance the accuracy and efficiency of tomato leaf disease detection.

Similarly, Guan *et al.* (2023) developed a lightweight model based on the EfficientNetV2 architecture, leveraging a dynamic learning rate decay strategy and transfer learning to achieve high recognition accuracy while maintaining a compact size, making it suitable for mobile and embedded devices. Further advancements have targeted specific technical challenges. Schuler *et al.* (2022) proposed a model that processes color information by separating color and grayscale components in the CIE Lab color space. This approach reduced computational complexity and improved classification accuracy, offering a novel perspective on handling color-sensitive tasks. Kiratiratanapruk *et al.* (2022) addressed issues caused by image size variation by using image tiling techniques and estimating leaf width, providing an effective solution for practical detection scenarios. Ahmed *et al.* (2022) applied data augmentation techniques to diversify training samples and improve model generalization, while Yao *et al.* (2024) adopted a multi-output structure to jointly optimize plant recognition and disease classification, further enhancing model performance. Additionally, transfer learning has proven effective in overcoming data limitations. Osouli *et al.* (2022) combined MobileNetV2 and Inception networks with transfer learning to develop a deep learning-based solution for corn disease recognition, achieving significant improvements in accuracy and efficiency despite limited data. While previous methods primarily relied on CNN-based architectures, recent advancements have started exploring the potential of transformer models (Zuo *et al.*, 2022; Yu Z *et al.*, 2024), which offer enhanced feature representation and global context understanding, addressing the limitations of CNNs in handling complex details and relationships (Guo *et al.*, 2024). A self-supervised transformer-based pre-training method using latent semantic masking auto-encoder (LSMAE) is proposed to improve pest and disease classification (Liu *et al.*, 2022). Experiments on public datasets show that this method outperforms CNN-based models, achieving higher accuracy, with 76.99% accuracy CPB. Liu *et al.* (2025) proposed an end-to-end pest detection method that combines feature representation compensation (FRC) and regional pale-shaped self-attention (RPSA) to address challenges in pest detection. Experiments on FPD datasets show that the method outperforms state-of-the-art CNN-based methods, achieving mAP scores that surpass previous methods by 5.7%. These studies collectively highlight the potential of CNN-based architectures to address various challenges in agricultural pest and disease detection, offering robust and efficient solutions for sustainable agricultural production.

In summary, these contributions pave the way for future innovations, demonstrating how deep learning can transform the field of precision agriculture. These methods have been compared and summarized in Table 4, providing a comprehensive overview of their performance and applications.

Current representative multimodal methods

To enhance the robustness and accuracy of pest and disease recognition, multimodal approaches have been increasingly explored in recent years. This section presents representative works in this area, with a focus on their design strategies and performance. Wang C *et al.* (2021) introduced ITC-Net, a few-shot learning model that combines disease images with textual descriptions, achieving remarkable results in recognizing plant diseases with limited samples. By leveraging both image and text modalities, ITC-Net outperforms traditional single-modality approaches, demonstrating the power of multimodal learning in small sample scenarios. Similarly, Zhou *et al.* (2021) proposed ITK-Net, which enhances disease recognition by incorporating knowledge graphs along with image and text data. The synergy of these modalities not only improves classification accuracy but also provides deeper semantic insights into the disease recognition process. In the realm of crop disease recognition, Cao *et al.* (2023) developed ITLMLP, a multimodal model that uses contrast learning to extract features from both image and text, achieving 94.84% accuracy in cucumber disease classification. This approach highlights the effectiveness of contrastive learning in small sample conditions. Meanwhile, Yu K *et al.* (2024) proposed ITFNet, which integrates attention-driven multimodal fusion for agricultural pest identification. Their model demonstrated improved performance by focusing on key visual and textual features, underscoring the importance of attention mechanisms in optimizing model performance. For plant disease description, Liang *et al.* (2024) introduced BLIP-DP, which dynamically generates cue words based on image content. This method significantly improves the accuracy of disease image description, enabling more precise pest and disease management. Additionally, VLCD was proposed by Zhou *et al.* (2024), a visual-language model that integrates image and text data for crop disease classification with few samples. The inclusion of attention mechanisms further enhanced the model's accuracy, emphasizing the growing role of multimodal fusion in disease classification.

To tackle pest recognition with limited samples, MMAE was introduced by

Zhang *et al.* (2025), a multimodal masked autoencoder model that combines image data with textual features. This model achieved a 98.12% recognition accuracy, outperforming traditional methods, and demonstrated the effectiveness of self-supervised learning techniques in small sample conditions. In the context of rice disease diagnosis, Patil and Kumar (2022) proposed Rice-Fusion, a multimodal fusion framework that integrates agrometeorological data with rice image data. This approach improved diagnostic accuracy to 95.31%, highlighting the potential of combining environmental sensor data with visual data for more reliable crop disease recognition. These methods, each leveraging unique strategies and techniques, are further compared and summarized in Table 5, providing a comprehensive overview of the current advancements in multimodal crop disease and pest recognition. In addition, we have also validated several general-purpose multimodal methods on the Agriculture-Vision dataset to further explore their applicability in agricultural scenarios (Wang *et al.*, 2022; Li *et al.*, 2023; Zhao *et al.*, 2023; Yang *et al.*, 2025).

Field deployments and technological developments in pest and disease recognition

Recent field deployments of pest and disease recognition technologies have demonstrated substantial benefits in real-world farm environments. In Kenya, farmers using the *Plant Village* app were able to identify and manage fall armyworm infestations precisely, avoiding unnecessary pesticide use. At IIT Kharagpur in India, an agricultural robot equipped with image-analysis modules can detect pests and diseases on leaves and spray insecticides automatically - reducing labor, minimizing health risks, and maintaining crop yields. In wheat and sorghum fields, real-time semantic segmentation models have successfully pinpointed aphid clusters, improving treatment accuracy and reducing pesticide waste. Moreover, in Argentina, a digital agriculture platform that integrates variable-rate spraying and soil mapping led to a 54% increase in peanut yields, highlighting how early detection and prevention of pests and diseases significantly contribute to crop production and resource efficiency. These real-world cases illustrate how combining image recognition, deep learning, and robot/drone systems can significantly enhance the effectiveness, sustainability, and economic outcomes of pest and disease management.

While these technologies demonstrate promising results in real-world deployments, their effectiveness is still subject to various environmental and hardware-related constraints. For example, fluctuating lighting conditions in open fields -such as shadows, backlighting, or low-light environments- can negatively affect image clarity and recognition performance. Moreover, many recognition models require substantial computational resources, which poses a challenge when deploying them on edge devices with limited memory and processing power. These practical limitations highlight the need for algorithms that are not only accurate in laboratory settings but also robust, lightweight, and adaptable to dynamic field conditions.

Current challenges and future directions

Current challenges

1) Data quality and availability: agricultural pest and disease detection rely on data from diverse sources, such as images, hyperspectral sensors, weather stations, and soil sensors. However, integrating these data types is challenging due to differences in formats, scales, and resolutions. For example, image data may require pixel-level analysis, while weather data is time-series-based. Additionally, the lack of high-quality, labeled datasets for training robust models limits the development of accurate and generalizable solutions. This heterogeneity complicates the creation of unified systems capable of leveraging multimodal data effectively.

2) Knowledge accessibility: one of the most pressing issues in agricultural AI is bridging the gap between complex technical outputs and practical, actionable insights for farmers. Many AI systems generate highly accurate predictions but fail to communicate these results in a way that is understandable and useful to non-experts. For instance, a model might identify a specific pest with high confidence but struggle to explain the reasoning behind its diagnosis or provide clear recommendations for treatment. This lack of interpretability and usability hinders the adoption of AI technologies in real-world farming scenarios.

3) Edge deployment: deploying AI models on edge devices, such as drones, smartphones, or IoT sensors, is essential for real-time pest and disease detection in the field. However, these devices often have limited computational power, memory, and energy resources, making it difficult to run complex models efficiently. Achieving real-time performance while maintaining high accuracy in dynamic and unpredictable field conditions is another major hurdle.

Additionally, optimizing models for energy efficiency to extend the battery life of edge devices remains a critical but unresolved challenge, especially in remote or resource-constrained agricultural settings.

Future directions

1) Multimodal data integration: future research should focus on developing efficient multimodal data fusion frameworks, integrating multisource heterogeneous data such as images, spectral data, and meteorological information through cross-modal representation learning and knowledge distillation techniques. Adaptive neural network architectures should be employed to address spatiotemporal scale discrepancies, while few-shot learning and self-supervised pretraining strategies should be explored to mitigate the scarcity of annotated data. Additionally, a privacy-preserving data collaboration platform based on federated learning should be constructed, integrating generative adversarial networks and physical models to enhance data synthesis capabilities. Finally, a human-machine collaborative annotation system should be integrated with domain knowledge graphs to form an interpretable agricultural decision support system, promoting a paradigm shift from single-modal analysis to multisource information collaborative perception.

2) Large language models for agricultural intelligence: LLMs, such as GPT (Floridi and Chiriatti, 2020) and DeepSeek (Liu A *et al.*, 2024), offer transformative potential for agriculture. These models can be fine-tuned to provide natural language explanations of pest and disease detection results, generate actionable recommendations for farmers, and even assist in knowledge dissemination by summarizing research papers or translating technical content into local languages. Additionally, LLMs can power conversational AI systems, enabling farmers to interact with detection tools using simple, intuitive queries.

3) Edge computing and real-time solutions: developing lightweight and efficient AI model architectures should be the primary focus of future research, utilizing techniques such as neural network pruning, quantization, and knowledge distillation to compress model size while ensuring high detection accuracy, thereby meeting the limited computational and storage resource requirements of edge devices. Building on this, further exploration of hybrid inference frameworks that combine edge and cloud computing can offload some computation-intensive tasks to the cloud, alleviating the burden on edge devices and improving the overall system's responsiveness. To address dynamic and unpredictable field conditions, adaptive model optimization algorithms need to be developed, enabling models to dynamically adjust computational complexity based on real-time environmental conditions, ensuring performance while reducing energy consumption. Additionally, the development of low-power hardware accelerators, such as dedicated AI chips or FPGAs, will be crucial for enhancing the computational efficiency of edge devices, coupled with novel energy harvesting technologies like solar or kinetic energy collection to significantly extend device battery life in remote agricultural settings. Finally, exploring collaborative computing mechanisms among edge devices and intelligent scheduling algorithms to optimize resource sharing and task allocation will provide critical support for achieving efficient, energy-saving, and sustainable real-time pest and disease detection. Comprehensive breakthroughs in these directions will drive the deep application of edge AI in agriculture, laying a solid foundation for the advancement of precision agriculture.

Conclusions

In this paper, we presented a comprehensive review of pest and disease recognition in agriculture, covering traditional machine learning, deep learning, and recent multimodal approaches. We summarized benchmark datasets, evaluation metrics, and representative algorithms, providing a holistic understanding of current developments. Despite notable progress, several technical bottlenecks remain that hinder real-world deployment. These include the limited computational capacity of edge devices, which constrains the application of complex models in the field; difficulties in acquiring high-quality, well-annotated, and diverse data; and challenges in effectively integrating multimodal inputs with varying formats and resolutions. Furthermore, the lack of model interpretability and user-friendly interfaces limits accessibility for non-expert users. Addressing these issues is essential for enhancing model performance, robustness, and usability. We hope this review inspires further research toward practical, intelligent, and scalable solutions for agricultural pest and disease management.

Acknowledgments This work was supported by the General project of National Nature Science Foundation of China (62176106), the Project of Faculty of Agricultural Engineering of Jiangsu University (NGXB20240101), the Special Scientific Research Project of School of Emergency Management of Jiangsu University (KY-A-01), the MTRAC Grant for Advanced Computing Technologles, the Jiangsu key research and development plan (industry foresight and key core technology, BE2020036) and the Key Project of Natlonal Nature Science Foundation of China (U1836220).

References

Ahmed, S., Hasan, M.B., Ahmed, T., Sony, M.R.K., Kabir, M.H. 2022. Less is more: Lighter and faster deep neural architecture for tomato leaf disease classification. IEEE Access 10:68868–68884.

Ali, S., Hassan, M., Kim, J.Y., Farid, M.I., Sanaullah, M., Mufti, H. 2022. Ff-pca-lda: intelligent feature fusion based pca-lda classification system for plant leaf diseases. Appl. Sci. 12:3514.

Bainalwar, P.A., Borkar, S.M., Shambharkar, S.S., Moon, P.S. 2023. Intelligent system to analysis of plant diseases using machine learning techniques. Proc. 5th Int. Conf. Information Management and Machine Intelligence, Salem. pp. 150-154.

Bollis, E., de Almeida Maia, H., Pedrini, H., Avila, S. 2021. Weakly supervised attention-based models using activation maps for citrus mite and insect pest classification. Comput. Electron. Agr. 195:106839.

Bollis, E., Pedrini, H., Avila, S. 2020. Weakly supervised learning guided by activation mapping applied to a novel citrus pest benchmark. Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle. pp. 310-319.

Cao, Y., Chen, L., Yuan, Y., Sun, G. 2023. Cucumber disease recognition with small samples using image-text-label-based multi-modal language model. Comput. Electron. Agr. 211;107993.

Cheng, J., Sun, J., Shi, L., Dai, C. 2024. An effective method fusing electronic nose and fluorescence hyperspectral imaging for the detection of pork

freshness. Food Biosci. 59:103880.
Chithambarathanu, M., Jeyakumar, M. 2023. Survey on crop pest detection using deep learning and machine learning approaches. Multimed. Tools Appl. 82:42277-42310.
Chiu, M.T., Xu, X., Wei, Y., Huang, Z., Schwing, A.G., Brunner, R., *et al*. 2020. Agriculture-vision: A large aerial image database for agricultural pattern analysis. Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle. pp. 2825-2836.
Dai, G., Fan, J., Dewi, C. 2023. Itf-wpi: Image and text based cross-modal feature fusion model for wolfberry pest recognition. Comput. Electron. Agr. 212:108129.
De Silva, M., Brown, D. 2022. Plant disease detection using deep learning on natural environment images. Proc. Int. Conf. Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban. pp. 1-5.
De Silva, M., Brown, D. 2023. Plant disease detection using vision transformers on multispectral natural environment images. Proc. Int. Conf. Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban. pp. 1-6.
Deng, F., Mao, W., Zeng, Z., Zeng, H., Wei, B. 2022. Multiple diseases and pests detection based on federated learning and improved faster r-cnn. IEEE T. Instrumt. Meas. 71:3523811.
Deng, L., Wang, Y., Han, Z., Yu, R. 2018. Research on insect pest image detection and recognition based on bio-inspired methods. Biosyst. Eng. 169:139-148.
El-Mesery, H.S., Adelusi, O.A., Ghashi, S., Njobeh, P.B., Hu, Z., Kun, W. 2024. Effects of storage conditions and packaging materials on the postharvest quality of fresh Chinese tomatoes and the optimization of the tomatoes' physiochemical properties using machine learning techniques. LWT 201:116280.
Ferentinos, K.P. 2018. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agr. 145:311-318.
Floridi, L., Chiriatti, M. 2020. Gpt-3: Its nature, scope, limits, and consequences. Mind. Mach. 30:681-694.
Gomes, J.C., Borges, D.L. 2022. Insect pest image recognition: A few-shot machine learning approach including maturity stages classification. Agronomy 12:1733.
Guan, H., Fu, C., Zhang, G., Li, K., Wang, P., Zhu, Z. 2023. A lightweight model for efficient identification of plant diseases and pests based on deep learning. Front. Plant Sci. 14:1227011.
Guo, Z., Zou, Y., Sun, C., Jayan, H., Jiang, S., El-Seedi, H.R., Zou, X. 2024.

Nondestructive determination of edible quality and watercore degree of apples by portable vis/nir transmittance system combined with cars-cnn. J. Food Meas. Charact. 18:4058-4073.
He, J., Chen, K., Pan, X., Zhai, J., Lin, X. 2023. Advanced biosensing technologies for monitoring of agriculture pests and diseases: A review. J. Semiconduct. 44:023104.
Huang, W., Zhu, W., Ma, C., Guo, Y., Chen, C. 2018. Identification of group-housed pigs based on gabor and local binary pattern features. Biosyst. Eng. 166:90-100.
Hughes, D., Salathé, M. 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv:1511.08060.
Jouini, O., Aoueileyine, M.O.E., Sethom, K., Yazidi, A. 2024. Wheat leaf disease detection: A lightweight approach with shallow cnn based feature refinement. AgriEngineering 6:2001-2022.
Junaid, M., Gokce, A. 2024. Global agricultural losses and their causes. Bull. Biol. All.S ci. Res. 2024:66.
Kale, M., Shitole, M. 2021. Analysis of crop disease detection with svm, knn and random forest classification. Inf. Technol. Ind. 9:364-372.
Kasinathan, T., Uyyala, S.R. 2021. Machine learning ensemble with image processing for pest identification and classification in field crops. Neural Comput. Appl. 33:7491-7504.
Kerkech, M., Hafiane, A., Canals, R. 2020. Vine disease detection in uav multispectral images using optimized image registration and deep learning segmentation approach. Comput. Electron. Agr. 174:105446.
Kethineni, K., Pradeepini, G. 2023. Identification of leaf disease using machine learning algorithm for improving the agricultural system. J. Adv. Inf. Technol. 14:122-129.
Khulal, U., Zhao, J., Hu, W., Chen, Q. 2016. Nondestructive quantifying total volatile basic nitrogen (tvb-n) content in chicken using hyperspectral imaging (hsi) technique combined with different data dimension reduction algorithms. Food Chem. 197:1191-1199.
Kiratiratanapruk, K., Temniranrat, P., Sinthupinyo, W., Marukatat, S., Patarapuwadol, S. 2022. Automatic detection of rice disease in images of various leaf sizes. IET Image Process. 19:e13301.
Kumar K.K., E, K. 2022. Detection of rice plant disease using AdaBoostSVM classifier. Agronomy J. 114:2213-2229.
Li, C.C., Yin, L.W., Chen, D. 2014. Study on plant pest images identification based on lifting wavelet transform. Appl. Mech. Mat. 687:3640-3643.
Li, H., Xu, T., Wu, X., Lu, J., Kittler, J. 2023. Lrrnet: A novel representation learning guided fusion network for infrared and visible images. IEEE T.

Pattern Anal. 45:11040–11052.
Liang, F., Huang, Z., Wang, W., He, Z., En, Q. 2024. Dynamic text prompt joint multimodal features for accurate plant disease image captioning. Vis. Comput. 41:5405-5019.
Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., *et al.* 2024. Deepseek-v3 technical report. arXiv:2412.19437.
Liu, H., Mao, Q., Dong, M., Zhan, Y. 2024. Infrared-visible image fusion using dual-branch auto-encoder with invertible high-frequency encoding. IEEE T. Circ. Syst. Vid. 35:2675-2688.
Liu, H., Zhan, Y., Sun, J., Mao, Q., Wu, T. 2025. A transformer-based model with feature compensation and local information enhancement for end-to-end pest detection. Comput. Electron. Agr. 231:109920.
Liu, H., Zhan, Y., Xia, H., Mao, Q., Tan, Y. 2022. Self-supervised transformer-based pre-training method using latent semantic masking auto-encoder for pest and disease classification. Comput. Electron. Agr. 203:107448.
Liu, J., Abbas, I., Noor, R.S. 2021. Development of deep learning-based variable rate agrochemical spraying system for targeted weeds control in strawberry crop. Agronomy 11:1480.
Liu, J., Wang, X. 2024. A multimodal framework for pepper diseases and pests detection. Sci. Rep. 14:28973.
Lu, B., Jun, S., Ning, Y., Xiaohong, W., Xin, Z. 2021. Identification of tea white star disease and anthrax based on hyperspectral image information. J. Food Proc. Eng. 44:e13584.
Mensah, P.K., Akoto-Adjepong, V., Adu, K., Ayidzoe, M.A., Bediako, E.A., Nyarko-Boateng, O. *et al.* 2023. Ccmt: Dataset for crop pest and disease detection. Data Brief 49:109306.
Mohanty, S.P., Hughes, D.P., Salathé, M. 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7:1419.
Nanni, L., Manfe, A., Maguolo, G., Lumini, A., Brahnam, S. 2021. High performing ensemble of convolutional neural networks for insect pest image detection. arXiv:2108.12539v1.
Navaneethan, S., Sampath, J.L., Kiran, S.S. 2023. Development of a multi-sensor fusion framework for early detection and monitoring of corn plant diseases. Proc. 2nd Int. Conf. on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai. pp. 856-861.
Nti, I.K., Eric, G., Jonas, Y.S. 2017. Detection of plant leaf disease employing image processing and gaussian smoothing approach. Int. J. Comput. Appl. 162:20-25.
Osouli, S., Haghighi, B.B., Sadrossadat, E. 2022. An effective scheme for maize disease recognition based on deep networks. arXiv:2205.04234.
Patil, R.R., Kumar, S. 2022. Rice-fusion: A multimodality data fusion framework

for rice disease diagnosis. IEEE Access 10:5207-5222.

Pawara, P., Okafor, E., Schomaker, L., Wiering, M. 2017. Data augmentation for plant classification. Proc. Advanced Concepts for Intelligent Vision Systems: ACIVS 2017. Lecture Notes in Computer Science, vol 10617. Cham, Springer. pp. 615-626.

Pei, H., Sun, Y., Huang, H., Zhang, W., Sheng, J., Zhang, Z. 2022. Weed detection in maize fields by uav images based on crop row preprocessing and improved Yolov4. Agriculture 12:975.

Peng, Y., Zhao, S., Liu, J. 2021. Fused-deep-features based grape leaf disease diagnosis. Agronomy 11:2234.

Pujari, J.D., Yakkundimath, R., Byadgi, A.S. 2015. Image processing based detection of fungal diseases in plants. Procedia Comput. Sci. 46:1802–1808.

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D.P. 2017. Deep learning for image-based cassava disease detection. Front. Plant Sci. 8:1852.

Rehana, H., Ibrahim, M., Ali, M.H. 2023. Plant disease detection using region-based convolutional neural network. arXiv:2303.09063.

Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., Siegwart, R. 2017. weednet: dense semantic weed classification using multispectral images and mav for smart farming. IEEE Robot. Autom. Lett. 3:588–595.

Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., Nelson, A. 2019. The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol. 3:430-439.

Schuler, J.P.S., Romani, S., Abdel-Nasser, M., Rashwan, H., Puig, D. 2022. Color-aware two-branch dcnn for efficient plant disease classification. Mendel 28:55-62.

Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K. 2020. Deep feature based rice leaf disease identification using support vector machine. Comput. Electron. Agr. 175:105527.

Shi, Y., Huang, W., Luo, J., Huang, L., Zhou, X. 2017. Detection and discrimination of pests and diseases in winter wheat based on spectral indices and kernel discriminant analysis. Comput. Electron. Agr. 141:171-180.

Singh, P., Verma, A., Alex, J.S.R. 2021. Disease and pest infection detection in coconut tree through deep learning techniques. Comput. Electron. Agr. 182:105986.

Tahir, H.E., Xiaobo, Z., Tinting, S., Jiyong, S., Mariod, A.A. 2016. Near-infrared (nir) spectroscopy for rapid measurement of antioxidant properties and discrimination of sudanese honeys from different botanical origin. Food Anal. Meth. 9:2631-2641.

Tian, Y., Sun, J., Zhou, X., Yao, K., Tang, N. 2022. Detection of soluble solid content in apples based on hyperspectral technology combined with deep learning algorithm. J. Food Process. Preserv. 46:e16414.

Wang, C., Zhou, J., Zhao, C., Li, J., Teng, G., Wu, H. 2021. Few-shot vegetable disease recognition model based on image text collaborative representation learning. Comput. Electron. Agr. 184:106098.

Wang, D., Liu, J., Fan, X., Liu, R. 2022. Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration. arXiv:2205.11876.

Wang, J., Gao, Z., Zhang, Y., Zhou, J., Wu, J., Li, P. 2021. Real-time detection and location of potted flowers based on a zed camera and a yolo v4-tiny deep learning algorithm. Horticulturae 8:21.

Wang, X., Xiao, Z., Deng, Z. 2025. Swin attention augmented residual network: a fine-grained pest image recognition method. Front. Plant Sci. 16:1619551.

Wang, Y., Zhang, X., Ma, G., Du, X., Shaheen, N., Mao, H. 2021. Recognition of weeds at asparagus fields using multi-feature fusion and backpropagation neural network. Int. J. Agr. Biol. Engin. 14:19-198.

Wang, Z., Wang, R.F., Wang, M., Lai, T., Zhang, M. 2024. Self-supervised transformer-based pre-training method with general plant infection dataset. Proc. Advanced Concepts for Intelligent Vision Systems: PRCV 2024. Lecture Notes in Computer Science, vol 15032., Singapore, Springer. pp 189–202.

Wei, L., Yang, H., Niu, Y., Zhang, Y., Xu, L., Chai, X. 2023. Wheat biomass, yield, and straw-grain ratio estimation from multi-temporal uav-based rgb and multispectral images. Biosyst. Eng. 234:187-205.

Wu, H., Li, X., Lu, H., Tong, L., Kang, S. 2023. Crop acreage planning for economy-resource-efficiency coordination: Grey information entropy based uncertain model. Agr. Water Manage. 289:108557.

Wu, X., Zhan, C., Lai, Y.K., Cheng, M.M., Yang, J. 2019. Ip102: A large-scale benchmark dataset for insect pest recognition. Proc. IEEE/CVF conference on computer vision and pattern recognition (CVPR), Long Beach. pp. 8787–8796.

Xu, S., Xu, X., Zhu, Q., Meng, Y., Yang, G., Feng, H., *et al.* 2023. Monitoring leaf nitrogen content in rice based on information fusion of multi-sensor imagery from UAV. Prec. Agr. 24:2327-2349.

Yang, B., Jiang, Z., Pan, D., Yu, H., Gui, G., Gui, W. 2025. Lfdt-fusion: A latent feature-guided diffusion transformer model for general image fusion. Inf. Fusion 113:102639.

Yang, N., Chang, K., Dong, S., Tang, J., Wang, A., Huang, R., Jia, Y. 2022. Rapid image detection and recognition of rice false smut based on mobile smart devices with anti-light features from cloud database. Biosyst. Eng. 218:229-244.

Yang, N., Qian, Y., EL-Mesery, H.S., Zhang, R., Wang, A., Tang, J. 2019a. Rapid detection of rice disease using microscopy image identification based on the synergistic judgment of texture and shape features and decision tree–confusion matrix method. J. Sci. Food Agr. 99:6589–6600.

Yang, N., Yu, J., Wang, A., Tang, J., Zhang, R., Xie, L., *et al.* 2020. A rapid rice blast detection and identification method based on crop disease spores' diffraction fingerprint texture. J. Sci. Food Agr. 100:3608–3621.

Yang, N., Yuan, M., Wang, P., Zhang, R., Sun, J., Mao, H. 2019). Tea diseases detection based on fast infrared thermal image processing technology. J. Sci. Food Agr. 99:3459–3466.

Yao, J., Tran, S.N., Garg, S., Sawyer, S. 2024. Deep learning for plant identification and disease classification from leaf images: multi-prediction approaches. ACM Comput. Surv. 56:1-37.

Yao, K., Sun, J., Tang, N., Xu, M., Cao, Y., Fu, L., *et al.* 2021. Nondestructive detection for panax notoginseng powder grades based on hyperspectral imaging technology combined with cars-pca and mpa-lssvm. J. Food Proc. Eng. 44:e13718.

Yu, K., Xu, W., Wu, Y. 2024. Attention driven multimodal fusion: Application of manual templates in feature aggregation. Proc. 3rd Int. Conf. on Computer, Artificial Intelligence and Control Engineering. pp. 878–886.

Yu, Z., Guo, Y., Zhang, L., Ding, Y., Zhang, G., Zhang, D. 2024. Improved lightweight zero-reference deep curve estimation low-light enhancement algorithm for night-time cow detection. Agriculture 14:1003.

Zhang, H., Mahunu, G.K., Castoria, R., Apaliya, M.T., Yang, Q. 2017. Augmentation of biocontrol agents with physical methods against postharvest diseases of fruits and vegetables. Trends Food Sci. Technol. 69:36–45.

Zhang, L., Wang, A., Zhang, H., Zhu, Q., Zhang, H., Sun, W., Niu, Y. 2024. Estimating leaf chlorophyll content of winter wheat from uav multispectral images using machine learning algorithms under different species, growth stages, and nitrogen stress conditions. Agriculture 14:1064.

Zhang, X., Bian, F., Wang, Y., Hu, L., Yang, N., Mao, H. 2022. A method for capture and detection of crop airborne disease spores based on. microfluidic chips and micro raman spectroscopy. Foods 11:3462.Zhang, Y., Chen, L., Yuan, Y. 2025. Few-shot agricultural pest recognition based on multimodal masked autoencoder. Crop Prot. 187:106993.

Zhang, Z., Lu, Y., Yang, M., Wang, G., Zhao, Y., Hu, Y. 2024. Optimal training strategy for high-performance detection model of multi-cultivar tea shoots based on deep learning methods. Sci. Horticult. 328:112949.

Zhao, S., Peng, Y., Liu, J., Wu, S. 2021. Tomato leaf disease diagnosis based on improved convolution neural network by attention module. Agriculture 11:651.

Zhao, Z., Bai, H., Zhang, J., Zhang, Y., Zhang, K., Xu, S., *et al*. 2023. Equivariant multi-modality image fusion. Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Seattle. pp. 25912–25921.

Zhou, J., Li, J., Wang, C., Wu, H., Zhao, C., Teng, G. 2021. Crop disease identification and interpretation method based on multimodal deep learning. Comput. Electron. Agr. 189:106408.

Zhou, X., Sun, J., Tian, Y., Lu, B., Hang, Y., Chen, Q. 2020. Hyperspectral technique combined with deep learning algorithm for detection of compound heavy metals in lettuce. Food Chem. 321:126503.

Zhou, Y., Yan, H., Ding, K., Cai, T., Zhang, Y. 2024. Few-shot image classification of crop diseases based on vision–language models. Sensors (Basel) 24:6109.

Zhu, H., Wang, D., Wei, Y., Zhang, X., Li, L. 2024. Combining transfer learning and ensemble algorithms for improved citrus leaf disease classification. Agriculture 14:1549.

Zhu, J., Cai, J., Sun, B., Xu, Y., Lu, F., Ma, H. *et al*. 2023. Inspection and classification of wheat quality using image processing. Qual. Assur. Saf. Crop. 15:43–54.

Zhu, W., Feng, Z., Dai, S., Zhang, P., Wei, X. 2022a. Using uav multispectral remote sensing with appropriate spatial resolution and machine learning to monitor wheat scab. Agriculture 12:1785.

Zhu, W., Sun, J., Wang, S., Shen, J., Yang, K., Zhou, X. 2022b. Identifying field crop diseases using transformer-embedded convolutional neural network. Agriculture 12:1083.

Zuo, X., Chu, J., Shen, J., Sun, J. 2022. Multi-granularity feature aggregation with self-attention and spatial reasoning for fine-grained crop disease classification. Agriculture 12:1499.

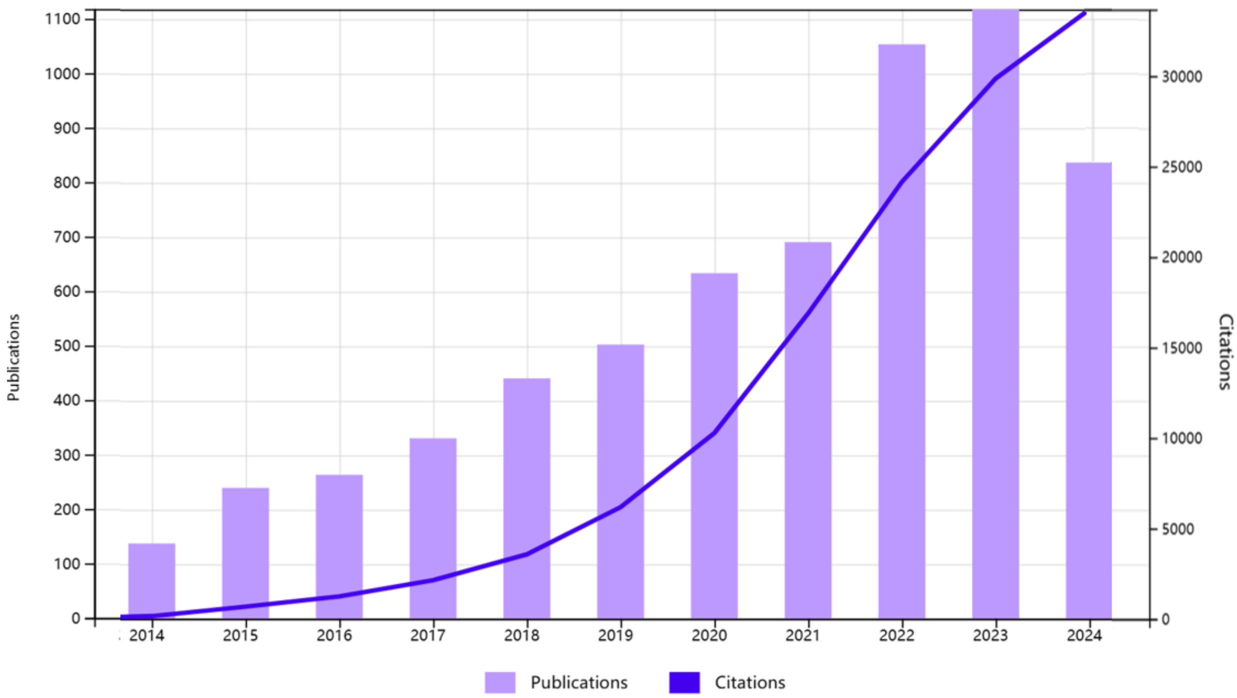

Figure 1. Trends in publications and citations on AI-based pest and disease recognition from 2014 to 2024. Data is from WOS advanced search.

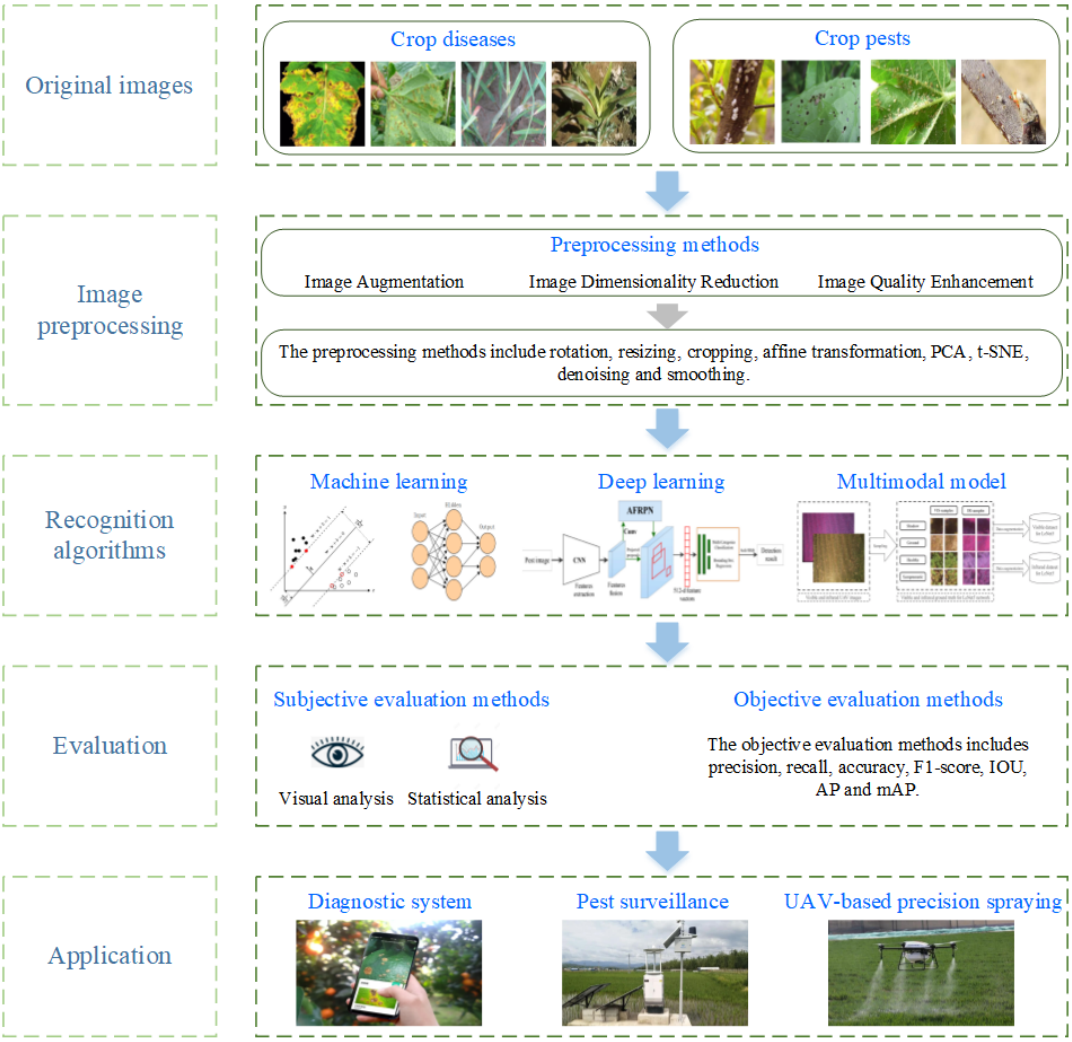

Figure 2. Specific steps of pest and disease recognition in agriculture.

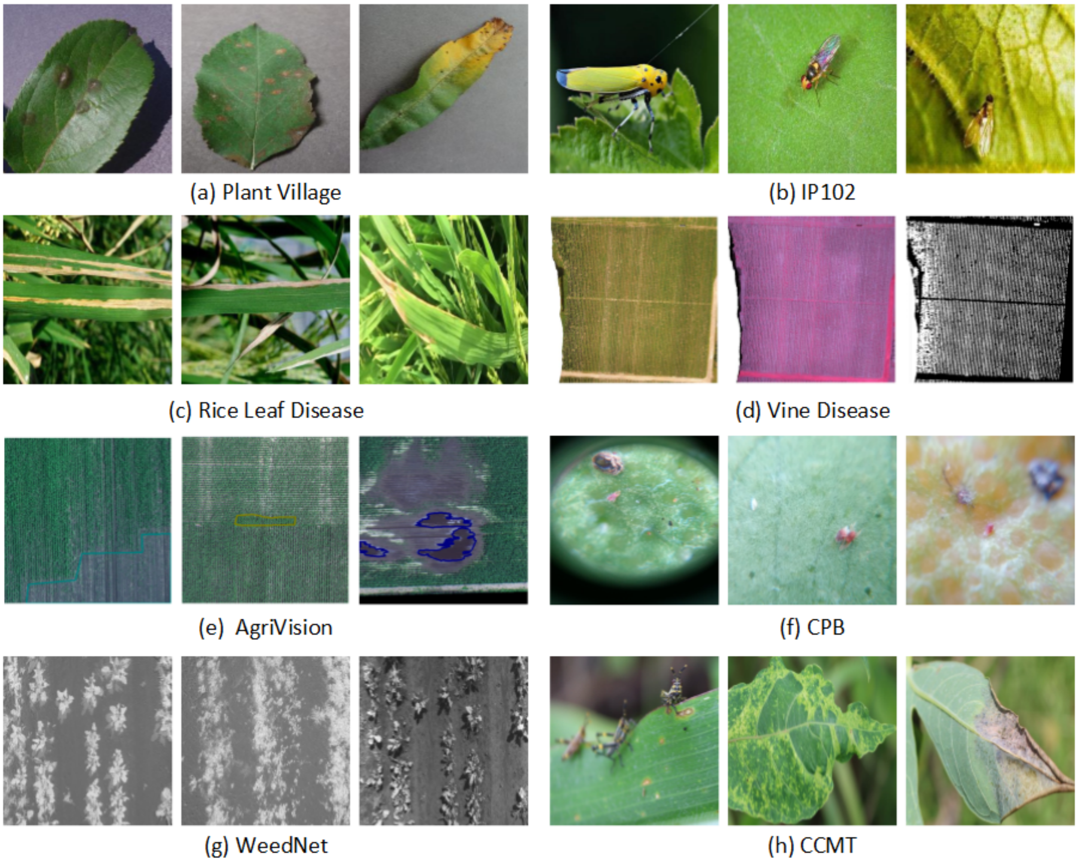

Figure 3. Specific steps of pest and disease recognition in agriculture.

Table 1. Benchmark datasets for pest and disease recognition.

References	Dataset	Type of data	Description	Source
Hughes *et al.* (2015)	Plant Village	RGB images	The Plant Village dataset consists of 38 categories, covering crops such as apples, blueberries, corn, grapes, and oranges.	Available
Wu *et al.* (2019)	IP102	RGB images	Insect pest images collected and annotated from the internet. It contains 102 insect pest categories at different stages, with 75,222 images total: 45,095 for training, 9,519 for validation, and 20,608 for testing.	Available
Sethy et al. (2020)	Rice Leaf Disease	RGB images	It covers four major rice leaf diseases, including bacterial leaf blight, blast disease, and brown spot, with a total of 5,132 augmented images for training and 800 images for testing.	Available
Kerkech *et al.* (2020)	Vine Disease	RGB and Thermal Images	This dataset includes visible and infrared images. The visible sensor detects chlorophyll presence, while the infrared sensor (850 nm) is sensitive to vegetation changes, aiding plant analysis.	--
Chiu *et al.* (2020)	AgriVision		AgriVision is a large-scale dataset for semantic segmentation of agricultural patterns, comprising 94,986 high-quality aerial images from 3,432 US farmlands.	Available
Bollis *et al.* (2020)	CPB	RGB images	The CPB consists of 10,816 images (1,200 × 1,200 pixels), divided into mite and negative classes. The images were collected using a Samsung Galaxy A5 with a 13 MP camera and magnification.	Available
Sa *et al.* (2017)	WeedNet	Multispectral images	WeedNet is a dataset designed for weed detection in agricultural fields. It contains a variety of images of both weeds and crops, collected under different environmental conditions.	Available
Mensah *et al.* (2023)	CCMT	RGB images	The CCMT dataset is a comprehensive collection of 24,881 raw color images spanning 22 classes, sourced from local farms in Ghana and focused on crop pests and diseases.	Available

Table 2. Confusion matrix.

		1 (positive)	0 (negative)
Predict	1 (positive)	TP (true positive)	FN (false negative)
	0 (negative)	FP (false positive)	TN (true negative)

Table 3. Summary of representative ML methods.

References	Datasets	Applications	Results or accuracy
Deng *et al*. (2018)	Self-collected dataset	Pest recognition	85.5%
Yang *et al*. (2019a)	Self-collected dataset	Rice disease classification	94.0%
Bainalwar *et al*. (2023)	Self-collected dataset	Plant disease detection	-
Kethineni and Pradeepini (2023)	Self-collected dataset	Rice leaf disease classification	98.8%
Gomes and Borges (2025)	IP-FSL dataset	Pest classification	Adult pest: 86.33% Early-stage pest: 87.91%
Kasinathan and Uyyala (2021)	Butterfly image dataset	Butterfly classification	92.3%

Table 4. Summary of representative DL methods.

References	Datasets	Applications	Results or accuracy
Ahmed *et al*. (2022)	Plant Village	Plant disease classification	99.3%
Osouli *et al*. (2022)	Plant Village	Plant disease classification	97.0%
Rehana *et al*. (2023)	Plant Village	Plant disease classification	96.3%
Yao *et al*. (2024)	Plant Village	Plant disease classification	99.6%
Bollis *et al*. (2021)	IP102	Pest classification	68.3%
Nanni *et al*. (2021)	IP102	Pest classification	73.5%
Liu *et al*. (2022)	IP102	Pest classification	74.7%
Guan *et al*. (2023)	IP102	Pest classification	64.4%
Wang *et al*. (2024)	IP102	Pest classification	76.2%
Wang *et al*. (2025)	IP102	Pest classification	78.8%

Table 5. Summary of representative multimodal methods.

References	Datasets	Applications	Results or accuracy
Wang C *et al*. (2021)	Self-collected dataset	Vegetable disease recognition	99.5%
Zhou *et al*. (2021)	Self-collected dataset	Crop disease identification	99.6%
Patil and Kumar (2022)	Self-collected dataset	Rice disease classification	95.3%
Cao *et al*. (2023)	IDADP	Cucumber disease recognition	94.8%
Zhang *et al*. (2025)	IDADP	Pest recognition	98.1%
Liang *et al*. (2024)	Plant Village	Plant disease caption	83.4%
Zhou *et al*. (2024)	Plant Village	Plant disease classification	87.3%
Wang *et al*. (2022)	AgriVision	Disease recognition	43.9%
Li *et al*. (2023)	AgriVision	Disease recognition	42.7%
Zhao *et al*. (2023)	AgriVision	Disease recognition	46.0%
Yang *et al*. (2025)	AgriVision	Disease recognition	46.8%