A Comprehensive Review of Deep Learning Techniques for Crop Pest Detection

Mr Mounesh Arkachari B[1], Maithri[2], Nischitha[3], Deepthi[4], Shreya[5]
Faculty, Department of Information Science and Engineering[1]
Students, Department of Information Science and Engineering[2,3,4,5]
Alva's Institute of Engineering and Technology
Mijar, Mangalore, Karnataka, India
Email: maithrikalkura11@gmail.com

***Abstract*—Crop pest infestation remains one of the most significant threats to global food security, causing substantial economic losses and reducing agricultural productivity worldwide. Traditional manual inspection methods are labor-intensive, time-consuming, and often ineffective for early detection, particularly in large-scale farming operations. In recent years, deep learning has emerged as a transformative technology for automated pest detection, offering high accuracy, real-time processing, and scalability. This paper presents a comprehensive review of deep learning approaches for agricultural pest detection, covering object detection architectures (YOLO, Faster R-CNN, SSD), classification models (ResNet, EfficientNet, Vision Trans-formers), multimodal frameworks combining visual and textual data, and segmentation-based methods. We analyze state-of-the-art contributions including YOLOv8-based lightweight models, CNN-Transformer hybrids, IoT-integrated detection systems, and explainable AI techniques for pest recognition. Key datasets including Pest24, IP102, and Agricultural Pests Dataset are examined. We identify critical challenges including small object detection, class imbalance, similar species discrimination, and real-world deployment constraints. Finally, we present future research directions including few-shot learning, edge deployment, multimodal fusion, and integrated pest management systems with pesticide recommendation capabilities. This review serves as a comprehensive reference for researchers and practitioners developing intelligent pest monitoring solutions for precision agriculture.**

***Index Terms*—Agricultural pest detection, deep learning, YOLOv8, object detection, convolutional neural networks, vision transformers, smart agriculture, precision farming, multimodal learning, IoT sensors.**

I. INTRODUCTION

Agriculture is the backbone of global food production and economic development, providing sustenance for over 8 billion people worldwide. However, crop pests pose a persistent and escalating threat to agricultural productivity, causing estimated annual losses of 20–40% of global crop yields, translating to economic damages of $220–$290 billion each year [4]. Insect pests alone account for approximately 38% of these losses, with major food crops potentially suffering yield reductions of up to 70% in the absence of effective crop protection measures [1]. The agricultural pest control market has expanded rapidly, exceeding $8.5 billion, driven by increasing food demand and climate change-exacerbated pest problems [3].

Traditional pest detection methods rely heavily on manual field scouting, where agricultural experts visually inspect crops or monitor physical traps [7]. While these approaches have been used for decades, they suffer from fundamental limitations: they are labor-intensive, time-consuming, subjective, and impractical for large-scale monitoring. Manual inspection is prone to human error, particularly when identifying visually similar pest species or detecting early-stage infestations. Furthermore, the delayed identification inherent in manual methods often results in reactive rather than proactive pest management, leading to excessive pesticide application, environmental contamination, and the development of pest resistance [6].

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized computer vision applications in agriculture. Deep learning models can automatically learn hierarchical features from raw image data, eliminating the need for handcrafted feature engineering that limited traditional machine learning approaches [2]. These models have demonstrated remarkable performance in pest detection, classification, and localization tasks, often exceeding human-level accuracy while operating at speeds suitable for real-time monitoring.

This paper provides a comprehensive review of deep learning approaches for agricultural pest detection, with the following contributions:

1) A systematic analysis of object detection architectures including YOLO series (YOLOv4 through YOLOv11), Faster R-CNN, SSD, and their applications in pest detection.
2) An examination of classification-based approaches using ResNet, EfficientNet, Vision Transformers (ViT), and hybrid CNN-Transformer architectures.
3) A review of multimodal frameworks integrating visual data with textual descriptions, IoT sensor data, and temporal information.
4) Analysis of segmentation-based methods using Mask R-CNN and hierarchical masking techniques for precise pest

localization.

5) Identification of key challenges including small object detection, class imbalance, similar species discrimina-tion, and computational constraints for edge deployment.
6) Future directions including few-shot learning, explain-able AI, and integrated pest management systems with pesticide recommendation capabilities.

The remainder of this paper is organized as follows: Section II presents the background on deep learning fundamentals and evaluation metrics. Section III reviews related work and state-of-the-art pest detection systems. Section IV examines datasets and preprocessing techniques. Section V analyzes detection, classification, and multimodal approaches. Section VI discusses key challenges and limitations. Section VII presents future research directions, and Section VIII concludes the paper.

II. BACKGROUND

A. Convolutional Neural Networks for Pest Detection

Convolutional Neural Networks form the foundation of modern deep learning-based pest detection systems. A typical CNN architecture consists of alternating convolutional and pooling layers, followed by fully connected layers for clas-sification. The convolutional layers apply learnable filters to extract hierarchical features, from low-level edges and textures to high-level semantic representations of pest morphology [1].

Key architectural innovations that have proven particularly effective for pest detection include:

Residual Networks (ResNet): Introduced by He et al., ResNet addresses the vanishing gradient problem through skip connec-tions that allow gradients to flow directly through the network. ResNet-50 and ResNet-101 have been widely adopted as back-bones for pest detection, achieving classification accuracies of 99.40% on benchmark datasets [6]. The residual connections enable training of very deep networks that capture fine-grained morphological details critical for distinguishing similar pest species.

EfficientNet: This family of architectures uses compound scaling to uniformly scale network depth, width, and resolu-tion. EfficientNet-B0 through B4 variants have demonstrated excellent performance on pest classification tasks, balancing accuracy with computational efficiency [2]. The compound scaling approach is particularly well-suited for pest detection, where pests appear at varying scales and resolutions in field images.

MobileNet: Designed for mobile and edge devices, MobileNet employs depthwise separable convolutions to reduce param-eter count and computational cost. MobileNetV3 has been successfully integrated into Faster R-CNN architectures for pest detection, achieving 92.66% mAP while maintaining high inference speed suitable for smartphone-based citizen science applications [1].

Vision Transformers (ViT): Recently, transformer-based ar-chitectures have emerged as powerful alternatives to CNNs for image classification. ViT divides images into patches and applies self-attention mechanisms to capture global contextual relationships. Costa et al. [2] demonstrated that ViT achieves 95.72% accuracy on the Agricultural Pests Dataset, outper-forming CNN-based approaches without data augmentation.

B. Object Detection Architectures

ResNet architectures address vanishing gradient problems using skip connections. ResNet-50 and ResNet-101 are widely used for pest classification tasks due to their ability to learn deep feature representations [6].

C. EfficientNet

EfficientNet uses compound scaling to balance network depth, width, and image resolution. These models provide high accuracy with lower computational complexity [2].

D. Vision Transformers (ViT)

Pest24: Created by Wang et al., Pest24 is the largest publicly available dataset for multi-target agricultural pest detection. It contains 25,378 images covering 24 pest species selected from the 38 categories designated by the Chinese Ministry of Agriculture. Key characteristics include extremely small target scales (pest relative scales primarily distributed between 0 and 0.01), incomplete annotation (only target pests labeled, non-target pests unlabeled), significant class imbalance (Anomala corpulenta has 53,347 instances vs. Holotrichia oblita with only 108), and challenging conditions including pest adhesion, occlusion, and variable illumination [8].

IP102: Wu et al. introduced IP102 as a large-scale benchmark for insect pest recognition, containing over 75,000 images spanning 102 pest species. The dataset exhibits a long-tailed distribution, with some species well-represented while others have limited samples. IP102 is particularly valuable for eval-uating model performance on rare pest species and transfer learning effectiveness

[1].

Agricultural Pests Dataset: Available on Kaggle, this dataset contains 5,494 images of 12 pest classes including ants, bees, beetles, caterpillars, moths, and wasps. Images are sourced from Flickr and resized to a maximum dimension of 300 pixels, providing diverse backgrounds, poses, and lighting conditions representative of real-world scenarios [2].

E. Data Preprocessing and Augmentation

Effective preprocessing is critical for pest detection due to the high variability in field-captured images:

Normalization: Images are normalized to zero mean and unit variance using dataset-specific statistics to stabilize training and improve convergence [4].

Data Augmentation: Given the limited size of many pest datasets, augmentation is essential to prevent overfitting and improve generalization. Common augmentations include ran-dom horizontal/vertical flips, rotations, color jitter, scaling, cropping, and mosaic augmentation [8]. Costa et al. [2] systematically evaluated augmentation strategies, finding that while ViT did not benefit from augmentation, CNNs showed significant improvements with techniques including random erasing and color jitter.

Class Imbalance Handling: Weighted loss functions assign higher importance to minority classes, with class weights calculated inversely proportional to class frequencies [4].

Focal loss has also been effective for addressing foreground-background imbalance in pest detection [8].

Near-Duplicate Removal: Butera et al. [1] emphasized the importance of removing near-duplicate images that can bias evaluation. Using ResNet50 embeddings and L2 distance thresholding, they eliminated 5-7% of samples that could cause data leakage between training and test splits.

F. Dataset Description

Table I presents a small comparative summary of datasets used in agricultural pest detection research.

TABLE I
SMALL AGRICULTURAL PEST DETECTION DATASETS

Dataset	Images	Classes	Purpose
Pest24	25,378	24	Pest Detection
IP102	75,000+	102	Pest Classification
Agricultural Pests Dataset	5,494	12	Image Classification
Rice Pest Dataset	4,800	9	Rice Pest Monitoring

The datasets listed above are widely used for training and evaluating deep learning models in agricultural pest detection systems. Pest24 and IP102 are large-scale datasets containing multiple pest categories and challenging environmental condi-tions. Smaller datasets such as Rice Pest Dataset are mainly used for crop-specific monitoring applications.

From the literature survey, it is observed that large-scale datasets such as IP102 and Pest24 provide significant diversity in pest species and environmental conditions, making them highly suitable for training robust deep learning models. Smaller datasets such as Rice Pest Dataset are useful for domain-specific applications but may suffer from limited gen-eralization capability. The availability of annotated datasets plays a critical role in improving the performance of CNN, YOLO, and Transformer-based pest detection systems.

III. DEEP LEARNING APPROACHES FOR PEST DETECTION

A. Object Detection Methods

1) YOLO-Based Detection: The YOLO (You Only Look Once) family has emerged as the dominant approach for real-time pest detection due to its exceptional speed-accuracy trade-off. YOLOv8, the latest iteration before YOLOv9-11, intro-duces anchor-free detection, decoupled heads, and improved augmentation strategies.

Cen et al. [8] proposed YOLO-LCE, a lightweight YOLOv8-based model specifically designed for agricultural pest detection on the Pest24 dataset. The model incorporates three key innovations:

1) **Lightweight Complementary Residual (LCR) Mod-ule:** The LCR module employs a dual-branch structure where one branch uses average pooling to extract stable pest features while the other uses max pooling to extract discriminative features. The complementary features improve discrimination between target and visually similar non-target pests. When in-tegrated into the C2f module (forming C2f-LCR), YOLO-LCE achieved 63.9% mAP50, a 1.7 percentage point improvement over baseline YOLOv8n.

2) **Efficient Partial Convolution (EPConv):** EPConv im-proves upon PConv by using an asymmetric channel splitting strategy

(1:7 ratio) where the smaller branch processes features through standard convolution while the larger branch uses lightweight group convolution. This ensures complete feature utilization while reducing computational complexity. The in-corporation of average pooling-based shortcuts from ResNet-D preserves fine-grained details beneficial for small pest targets.

3) **Ghost Module in Detection Head:** The Ghost module generates additional feature maps through cheap depthwise convolution operations, reducing computational overhead. Ap-plied to the computationally intensive decoupled detection head, the Ghost module reduced GFLOPs by 20.6% with only a 0.3% mAP50 decrease.

Additional optimizations include WIoUv3 loss function with dynamic non-monotonic gradient allocation and co-sine annealing learning rate scheduling. YOLO-LCE achieves 63.9% mAP50 and 39.1% mAP50-95 while requiring only 1.69M parameters and 5.4 GFLOPs, making it suitable for deployment on resource-constrained edge devices.

Other YOLO variants for pest detection include YOLOv4-tiny (55.1% mAP50), YOLOv5n (57.8%), YOLOv7-tiny (59.2%), YOLOv10n (61.8%), and YOLOv11n (62.2%), demonstrating the rapid evolution and increasing sophistica-tion of the YOLO series [8].

2) Faster R-CNN and Two-Stage Detectors: Faster R-CNN remains a strong baseline for pest detection, particularly when high accuracy is prioritized over inference speed. Butera et al. [1] conducted extensive experiments with Faster R-CNN using VGG16, ResNet101, DenseNet169, and MobileNetV3 backbones. The MobileNetV3-FPN combination achieved the best trade-off with 92.66% mAP and 60.92 FPS on GPU, while also producing the lowest false positive count for Popillia japonica detection (12 false positives out of 530 negative samples, a 2.26% false positive ratio).

Wu et al. [7] proposed hierarchical mask R-CNN (HM-R-CNN) and threshold-based hierarchical mask R-CNN (TbHM-R-CNN) for pest detection on crop leaves. The TbHM-R-CNN model achieved 96.2% classification accuracy, 97.5% recall, and 0.982 F1-score, outperforming conventional CNNs and SVM-based approaches. The radial bisymmetric divergence (RBD) method was introduced to generate hierarchical masks that select informative image regions, reducing storage require-ments and improving detection efficiency.

B. Classification-Based Approaches

For applications where pest presence rather than precise localization is required, image classification models offer efficient solutions.

Costa et al. [2] evaluated four architectures (AlexNet, ResNet-50, EfficientNet-B4, ViT) on the Agricultural Pests Dataset. ViT (Vision Transformer) achieved the highest test accuracy of 95.72% without data augmentation, outperforming CNN-based approaches. The self-attention mechanism of ViT effectively captures global contextual relationships, which is particularly beneficial for discriminating visually similar pest species. Interestingly, ViT did not benefit from data augmen-tation, while ResNet-50 and EfficientNet showed significant improvements with augmentation strategies including random flips, rotations, and color jitter.

Dewi et al. [6] applied transfer learning with ResNet-50 and ResNet-101 for insect pest classification on a 10-class dataset. ResNet-50 achieved an exceptional 99.40% accuracy, 99.10% precision, and 99.10% recall, significantly outperforming pre-vious methods including LCP+SVM (85.5%) and TL AlexNet (93.84%). The success of ResNet-50 is attributed to its residual connections, which enable training of deep networks that cap-ture fine-grained morphological features distinguishing similar pest species.

C. CNN-Transformer Hybrid Architectures

Recent work has explored hybrid architectures combining CNN backbones with transformer encoders to leverage the strengths of both paradigms: CNNs excel at local feature extraction while transformers capture global dependencies.

Verma et al. [4] evaluated multiple hybrid architectures including Hybrid InceptionResNetV2, Hybrid ResNet50+CBAM, Hybrid EffNet+Transformer, and Hybrid EfficientNetV2-S+Transformer for multi-class pest classification on 19 species. The Hybrid EfficientNetV2-S+Transformer achieved the highest performance with 88.0% validation accuracy, 0.849 macro-F1, and 0.456 validation loss. The model combines EfficientNetV2-S's compound scaling and fused-MBConv blocks for efficient local feature extraction with a transformer encoder that models long-range dependencies across image patches. The CLS token pooled from transformer output provides a global representation that significantly improves discrimination of visually similar pests.

D. Multimodal Approaches

Recognizing that visual information alone may be insuf-ficient for reliable pest identification, several studies have explored multimodal integration.

Duan et al. [5] proposed a multimodal framework combin-ing tiny-BERT for natural language processing with R-CNN and ResNet-18 for image processing. Textual descriptions generated by LLAVA (Large Language and Vision Assistant) provide contextual information about pest behavior, environ-mental conditions, and crop damage patterns that complement visual morphology. Ensemble learning using weighted average achieved an AUC of 0.994, while linear regression achieved 0.977 and random forest achieved 0.985. The multimodal ap-proach effectively filters unimodal noise by cross-referencing consistent information across modalities.

Arunachalam et al. [3] developed an IoT-based pest detection system integrating Feature Pyramid Network (FPN) with Multi-Attention Fusion Vision Transformer and Adaptive LSTM (FPN-MAFViT-ALSTM). The system collects IoT sensor-based pest images and performs joint detection andclassification. An enhanced optimization algorithm (EIGBO-RE) tunes hyperparameters including learning rate and hidden neurons. The model achieved 96.72% accuracy on Dataset-1 and 96.69% on Dataset-3, outperforming existing methods including Faster-PestNet and PestNet.

E. Segmentation-Based Methods

Image segmentation provides precise pest localization at the pixel level, enabling accurate infestation severity assessment. Wu et al. [7] introduced radial bisymmetric divergence (RBD) to generate hierarchical masks that select informative image regions. The HM-R-CNN model combines mask R-CNN with RBD-based hierarchical masking, while TbHM-R-CNN adds a threshold-based fault-tolerant mechanism. Simple Linear Iterative Clustering (SLIC) performs final image segmentation based on selected features. The TbHM-R-CNN model achieved 96.2% accuracy, demonstrating that segmentation-based preprocessing significantly improves detection of leaf-dwelling pests while reducing storage requirements.

IV. CHALLENGES AND LIMITATIONS

A. Small Object Detection

The most significant challenge in pest detection is the extremely small size of many pest targets. On the Pest24 dataset, relative pest scales are primarily distributed between 0 and 0.01, meaning pests occupy very few pixels even in high-resolution images [8]. Small objects provide limited visual information, making feature extraction difficult. Standard detection architectures often struggle with small objects due to aggressive downsampling in backbone networks that discards fine-grained details. While methods such as feature pyramid networks (FPN) and multi-scale training partially address this issue, small pest detection remains an open problem requiring specialized architectures and training strategies.

B. Class Imbalance

Agricultural pest datasets exhibit severe long-tailed distributions, with common species having tens of thousands of instances while rare species have only hundreds or even dozens. This imbalance biases models toward dominant classes, degrading performance on rare but potentially equally important pest species [8]. For example, on Pest24, Holotrichia oblita has only 108 instances compared to 53,347 for Anomala corpulenta. Class imbalance handling techniques including weighted loss functions, oversampling, and synthetic data generation provide partial solutions but remain insufficient for extremely rare classes.

C. Similar Species Discrimination

Many pest species exhibit high visual similarity, making discrimination challenging even for human experts. Popillia japonica (harmful) is nearly indistinguishable from Cetonia aurata and Phyllopertha horticola (harmless) in images [1]. This similarity manifests in two forms: (1) similarity among target pests requiring fine-grained discrimination, and (2) similarity between target and non-target pests that are unlabeled in training data, causing false positives. Hybrid CNN-transformer architectures show promise for this challenge, as transformers capture subtle global patterns that CNNs may miss, but perfect discrimination remains elusive.

D. Real-World Deployment Constraints

Laboratory performance often fails to translate to field conditions due to:

- **Lighting variations:** Field images exhibit extreme lighting conditions including shadows, glare, and inconsistent illumination.
- **Background clutter:** Complex foliage, soil, and other plant parts create visual noise.
- **Pest occlusion and adhesion:** Overlapping pests and partial occlusion by leaves impede detection.
- **Imprecise annotations:** As noted by Cen et al. [8], bounding boxes in some datasets are loose and do not tightly fit pest boundaries, introducing label noise.
- **Computational constraints:** Edge deployment requires models with small memory footprint, low latency, and minimal energy consumption, conflicting with the high capacity needed for accurate detection.

E. Generalization Across Domains

Models trained on one dataset often fail to generalize to different crops, geographic regions, seasons, or imaging devices. The dataset shift problem is acute in agriculture due to the high variability in pest appearance across life stages, host plants, and environmental conditions. Transfer learning partially addresses this challenge, but collecting diverse, representative training data remains essential and costly.

V. FUTURE RESEARCH DIRECTIONS

A. Few-Shot and Zero-Shot Learning

Given the long-tailed distribution of pest species and the high cost of collecting labeled data for rare pests, few-shot learning approaches that can recognize novel pest species from only a few examples are critically needed. Meta-learning frameworks, prototypical networks, and metric learning have shown promise for few-shot classification and could be adapted for pest detection. Zero-shot learning using auxiliary information such as textual descriptions of pest characteristics offers an even more ambitious direction for recognizing pest species never seen during training.

B. Edge Deployment and Lightweight Models

The ultimate goal of many pest detection systems is deploy-ment on field devices with limited computational resources. Future research should focus on:

- **Model compression:** Quantization, pruning, and knowl-edge distillation to reduce model size and latency while preserving accuracy.
- **Neural architecture search (NAS):** Automated design of architectures optimized for specific hardware constraints.
- **On-device learning:** Continual learning approaches that adapt to new pest species or changing field conditions without cloud connectivity.
- **Energy-efficient inference:** Specialized hardware accel-erators and efficient operators for ultra-low-power oper-ation.

C. Advanced Multimodal Integration

Beyond visual-textual fusion, future systems should inte-grate:

- **Temporal information:** Video sequences capturing pest movement and behavior patterns.
- **Environmental data:** Temperature, humidity, and weather information that influence pest activity.
- **Spectral imaging:** Multispectral and hyperspectral data revealing pest stress invisible in RGB images.
- **Acoustic sensing:** Pest sounds for detection in dense foliage where visual occlusion is severe.
- **Pheromone trap data:** Quantitative trap catch informa-tion for population monitoring.

D. Explainable AI and Farmer Trust

For widespread adoption, pest detection systems must be interpretable and trustworthy. Guided GradCAM and other visualization techniques can highlight image regions used for pest identification, allowing experts to validate model decisions [1]. Future work should develop:

- **Uncertainty estimation:** Quantifying prediction confi-dence to flag ambiguous cases requiring human review.
- **Counterfactual explanations:** Demonstrating how small changes in input would alter predictions.
- **Natural language explanations:** Generating human-readable justifications for pest identifications.

E. Integrated Pest Management Systems

The ultimate vision is comprehensive pest management systems that not only detect pests but also:

- **Calculate infestation severity:** Percent of crop area affected or pest count per plant.
- **Recommend interventions:** Suggest appropriate pesti-cides, biological controls, or cultural practices based on pest species, infestation level, crop growth stage, and environmental conditions.
- **Predict pest outbreaks:** Time-series forecasting using historical data and environmental variables.
- **Enable precision spraying:** Integration with robotic sprayers for targeted, minimal pesticide application.

As outlined in the project presentations [**?**], [**?**], a complete system would detect pests using YOLOv8, calculate infesta-tion percentage to determine severity levels, and recommend suitable pesticides based on detected pest type, supporting the Integrated Pest Management (IPM) paradigm.

VI. CONCLUSION

This paper has provided a comprehensive review of deep learning approaches for agricultural pest detection, covering object detection architectures, classification models, multi-modal frameworks, and segmentation-based methods. The field has advanced rapidly, with YOLO-based detectors achiev-ing strong performance on challenging datasets like Pest24, Vision Transformers surpassing CNNs on classification tasks, and hybrid architectures combining the strengths of both paradigms. Key innovations including lightweight modules (LCR, EPConv), attention mechanisms, and optimization algo-rithms (EIGBO-RE) have significantly improved the accuracy-efficiency trade-off, making edge deployment increasingly feasible.

However, substantial challenges remain. Small object de-tection, class imbalance, similar species discrimination, and real-world deployment constraints continue to limit practical adoption. Future research must address these challenges through few-shot learning, multimodal integration, explainable AI, and comprehensive integrated pest management systems. The convergence of deep learning with IoT sensors, edge computing, and precision agriculture technologies promises to transform pest monitoring from reactive, labor-intensive manual inspection to proactive, automated, and intelligent systems that protect crop yields while minimizing environmental impact.

As agriculture faces the dual pressures of climate change and growing global population, automated pest detection will be an essential tool for sustainable food production. The research community must continue to develop robust, efficient, and trustworthy systems that empower farmers with actionable intelligence for pest management.

ACKNOWLEDGMENT

The authors would like to thank the Department of Information Science and Engineering, Alva's Institute of Engineering and Technology, for providing the facilities and support for this research. Special thanks to the faculty and staff for their continuous guidance and encouragement.

REFERENCES

[1] L. Butera, A. Ferrante, M. Jermini, M. Prevostini, and C. Alippi, "Precise agriculture: Effective deep learning strategies to detect pest insects," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 2, pp. 246–258, Feb. 2022.

[2] P. L. O. Costa, T. M. O. Costa, L. F. R. Moreira, L. H. F. P. Silva, and J. F. Mari, "Classification of agricultural pests through digital images using deep learning," 2025.

[3] R. Arunachalam, M. Jaishankar, A. Arora, P. Shanmugam, S. Venugopal, and T. P. Priyanka, "An efficient feature pyramid network with adaptive LSTM for pest detection and classification in IoT," *Sci. Rep.*, vol. 16, p. 4423, 2026.

[4] V. K. Verma, A. Kumar, V. Pareek, and Y. Kumar, "Transformer augmented hybrid deep learning for explainable multi class pest classification," *Sci. Rep.*, 2026.

[5] J. Duan, H. Ding, and S. Kim, "A multimodal approach for advanced pest detection and classification," 2025.

[6] C. Dewi, H. J. Christanto, and G. W. Dai, "Automated identification of insect pests: A deep transfer learning approach using ResNet," *Acadlore Trans. Mach. Learn.*, vol. 2, no. 4, pp. 194–203, 2023.

[7] X. Wu, Y. Liu, M. Xing, C. Yang, and S. Hong, "Image segmentation for pest detection of crop leaves by improvement of regional convolutional neural network," *Sci. Rep.*, vol. 14, p. 24160, 2024.

[8] X. Cen, S. Lu, and T. Qian, "YOLO-LCE: A lightweight YOLOv8 model for agricultural pest detection," *Agronomy*, vol. 15, p. 2022, 2025.