Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Design and Application of an Intelligent Plant Disease and Pest Recognition System for Landscape Architecture Based on Deep Learning

Na Li

How to cite: Li N. Design and Application of an Intelligent Plant Disease and Pest Recognition System for Landscape Architecture Based on Deep Learning. Textile & Leather Review. 2026; 9:6025-6038. https://doi.org/10.31881/TLR.2026.6025

How to link: https://doi.org/10.31881/TLR.2026.6025

Published: 5 June 2026

Design and Application of an Intelligent Plant Disease and Pest Recognition System for Landscape Architecture Based on Deep Learning

Na Li

Environmental Art and Design, Hebei Art & Design Academy, Baoding 071051, Hebei, China
18730263063@163.com

Article
https://doi.org/10.31881/TLR.2026.6025
Published 5 June 2026

ABSTRACT

Manual inspection of landscape pathology and textile fiber defects suffers from inherent subjective bias and suboptimal throughput. To bypass these bottlenecks, we propose the Ghost-Convolution Enlightened Vision Transformer (GeT). We constructed a novel hybrid neural network architecture, the Ghost-Convolution Enlightened Vision Transformer (GeT), which synergistically integrates the lightweight local feature extraction proficiency of Convolutional Neural Networks (CNN) with the global semantic modeling capabilities of Vision Transformers (ViT). Utilizing a newly established standard dataset (GLDP15k) comprising 15,000 heterogeneous field images, the model was subjected to rigorous hyperparameter optimization and ablation studies. Optimization on the GLDP15k dataset yielded a peak accuracy of 96.8% across 12 target classes, maintaining a Kappa-coefficient of 0.941. Constrained to 1.16 M parameters, the architecture executes at 5.5 ms per image (180 FPS) on edge hardware. A 6-month application at Yuexiu Park demonstrated a 3.2-fold improvement in detection efficiency and a 35% reduction in pesticide usage compared to manual inspections. This study not only elucidates the interpretability of hybrid attention mechanisms in phytopathology but also adapts these vision-based paradigms to the detection of microscopic anomalies in textile weaving patterns, providing a scalable and computationally efficient solution for both precision plant protection and industrial fabric defect inspection.

KEYWORDS

deep learning, vision transformer, ghost-convolution, plant diseases and pests, textile weaving patterns

INTRODUCTION

Plant diseases and pests in landscape architecture severely threaten the ecological quality of urban environments and the aesthetic value of garden landscapes, mirroring the detrimental impact that structural impurities and fiber degradation have on the quality of raw textile materials. Just as biological stressors

undermine plant health, these anomalies compromise the functional integrity and visual consistency of textile manufacturing, necessitating advanced monitoring to preserve both natural and industrial assets. Traditional disease and pest identification heavily relies on the visual judgment of professional technicians. This conventional paradigm is not only labor-intensive but also highly susceptible to misdiagnoses due to overlapping symptomatic expressions across different plant species and human fatigue [1-4]. Classic Convolutional Neural Networks (CNNs) have achieved high accuracies on controlled laboratory datasets like PlantVillage [5,6]. However, landscape plants present a unique challenge: they encompass diverse botanical species with highly varied symptom manifestations and are heavily impacted by heterogeneous urban environmental factors (e.g., erratic lighting, complex concrete backgrounds, and severe shadow occlusions) [7,8].

CNNs excel in extracting local textures via translation equivariance; they inherently lack the global receptive fields required to model spatially distributed symptom dependencies [9,10]. Vision Transformers (ViTs) resolve this spatial constraint through multi-head self-attention [11,12]. Pure ViT architectures, however, demand massive datasets to compensate for their absence of inductive bias, generating prohibitive computational overhead (FLOPs) incompatible with edge-devices [13,14].

To overcome these algorithmic limitations, this study proposes a lightweight hybrid architecture: the Ghost-Convolution Enlightened Vision Transformer (GeT). The main objectives are: (1) to design the GeT model that mathematically balances local feature induction with global attention mapping via cheap linear operations; (2) to optimize model hyper-parameters and validate its robustness on a complex field dataset (GLDP15k); (3) to interpret the model's decision-making process through attention map visualizations; and (4) to evaluate the system's field efficacy and ecological benefits through a large-scale urban park deployment.

MATERIALS AND METHODS

Dataset Acquisition and Augmentation

To address the lack of field-based landscape datasets, the Guangzhou Landscape Plant Disease and Pest Dataset (GLDP15k) was constructed. It contains 15,000 high-resolution images (1920 × 1080 pixels) collected via digital cameras and edge-computing surveillance nodes across varying microclimates. The dataset encompasses 12 classes: 7 diseases (Leaf Spot, Powdery Mildew, Anthracnose, Rust, Bacterial Wilt, Viral Diseases, and Healthy) and 5 pests (Aphids, Spider Mites, Longhorn Beetles, American White Moths, and Scale Insects) [15]. To eliminate majority-favor bias during training, class weighting and multi-scale data augmentation were applied. Let $X \in \mathbb{R}^{C\times H\times W}$ be the original image. Geometrical transformations (random affine translations,

horizontal/vertical flipping) and photometric distortions (Gaussian noise, HSV adjustments) were applied. Furthermore, the CutMix and MixUp regularization strategies were employed to synthesize high-fidelity training data, expanding minority classes to a minimum of 1,250 samples each to penalize overconfidence [16]. The dataset was partitioned via 5-fold cross-validation into training, validation, and test sets at an 8:1:1 ratio.

A baseline of 1,250 images per class initially seems insufficient against field-level morphological variance. Robust multi-scale augmentations (MixUp and CutMix) mitigated this constraint by synthesizing exponentially diverse virtual phenotypes. Transfer-learning supplied foundational feature representations, restricting the data volume required for convergence while preserving high generalization against complex urban backgrounds.

System Architecture Design

Field deployment necessitated a five-layer decoupled microservice architecture [17,18]:

1. Data Layer: Manages multi-source image acquisition and database storage (MySQL/MongoDB).
2. Algorithm Layer: Houses the GeT hybrid model. It utilizes TensorRT for model quantization (FP32 to INT8) for edge deployment without sacrificing mathematical precision [19].
3. Service Layer: Encapsulates DL inference as RESTful APIs, integrating Nginx load balancing.
4. Application Layer: Executes core business logic: real-time disease recognition and treatment recommendation generation via an expert rule system. The expert rule system operates via a structured mapping matrix validated by horticulturalists. Once the GeT model outputs a specific pathology class (e.g., Powdery Mildew) with a confidence score exceeding 0.85, the rule database dictates prescriptive interventions. Localized application of targeted fungicides (e.g., Triadimefon) explicitly replaces prophylactic broad-spectrum spraying. This rule-based mapping links visual recognition directly to practical field management.
5. Interaction Layer: Provides multi-platform access via Web portals and mobile applications [20].

Ghost-Convolution Enlightened Transformer (GeT) Architecture

Ghost-Convolution Backbone

Standard convolutions generate highly redundant feature maps, squandering computational resources. The Ghost module mitigates this by generating intrinsic feature maps via pointwise convolution and deriving "ghost" features through cheap linear operations.

For an input $X \in \mathbb{R}^{C\times H\times W}$, the intrinsic feature map $Y' \in \mathbb{R}^{m\times h'\times w'}$ is generated as:

$$Y'_{i,j} = \sum_{c=1}^{C} X_{i,j,c} * K_{c,m} + b \tag{1}$$

Then, a cheap depthwise linear operation Φ generates the ghost feature maps. $y_{i,j}$

$$y_{i,j} = \Phi_{i,j}(Y'_i), \quad \forall i = 1, ..., m; \quad j = 1, ..., s \tag{2}$$

The final output is the concatenation of Y' and y , yielding $n = m \times s$ feature maps, drastically reducing parameters. This backbone processes the image through 4 stages, progressively reducing spatial resolution while expanding channel dimensions.

Patch Embedding and Transformer Encoder

The output feature maps $X_G \in \mathbb{R}^{C_{oi} \times h_{oi} \times w_{oi}}$ from the backbone are processed by a 2D convolution with a stride P (patch down-sampling ratio) and flattened into N patches to form a sequence $X_F \in \mathbb{R}^{N \times D}$. A learnable class token X_{class} and positional embeddings E_{pos} are prepended:

$$z_0 = [X_{class}; X_F] + E_{pos} \tag{3}$$

The sequence z_0 is fed into an L -depth Transformer encoder. Each block consists of Layer Normalization (LN), Multi-Head Self-Attention (MHSA), and a Multi-Layer Perceptron (MLP) with GELU activations:

$$z'_l = MHSA(LN(z_{l-1})) + z_{l-1} \tag{4}$$

$$z_l = MLP(LN(z'_l)) + z'_l, \quad l = 1, ..., L \tag{5}$$

Loss Function and Model Optimization

To address hard-to-classify samples (e.g., distinguishing early bacterial wilt from physiological drought stress), an improved Focal Loss function was utilized:

$$L_{focal} = -\frac{1}{B} \sum_{b=1}^{B} \sum_{k=1}^{K} \alpha_k (1 - P_{b,k})^{\gamma} \log(P_{b,k}) \tag{6}$$

where $P_{b,k}$ is the predicted probability, α_k is the inverse class frequency weight, and $\gamma = 2.0$ is the focusing parameter that penalizes hard examples.

Experimental Setup and Evaluation Metrics

Training was executed on a workstation with Dual NVIDIA RTX 4090 GPUs using PyTorch 2.1. The AdamW optimizer was employed with a base learning rate of 1×10^{-3} , weight decay of 1×10^{-4} , and a cosine annealing learning rate scheduler over 150 epochs. Batch size was set to 64.

Model performance was evaluated using Accuracy, Precision, Recall, F1-score, and Inference Speed (Frames Per Second, FPS). Unbiasedness was quantified using the Cohen's Kappa coefficient (κ):

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{7}$$

where p_0 is the observed accuracy and p_e is the expected agreement probability.

Field Deployment and Evaluation Protocol

To validate the practical ecological benefits, a 6-month controlled field study was conducted at Yuexiu Park. The park was divided into two comparable zones. Zone A (Control) relied on traditional manual inspections and prophylactic calendar-based pesticide spraying. Zone B (Experimental) utilized our deployed edge-cloud system (UAV and surveillance cameras) to trigger targeted pesticide applications only when the system alerted to a confirmed infestation. The 3.2-fold efficiency improvement was calculated by comparing the average labor hours required to scan 1 hectare. The 35% pesticide reduction was measured by comparing the total chemical volume (liters) consumed in Zone B versus Zone A over the 6 months.

RESULTS

Architecture Validation and Baseline Comparison

To systematically evaluate the proposed framework, we first validated the model against the complex phenotypic variations present in our dataset (as shown in Figure 1). The overall edge-cloud collaborative deployment of the 5-layer system (Figure 2) and the forward-pass data flow of the GeT model (Figure 3) were computationally simulated. We compared the performance of our internal hybrid modules (Figure 4) against several State-Of-The-Art (SOTA) baseline models fine-tuned on GLDP15k (Table 1).

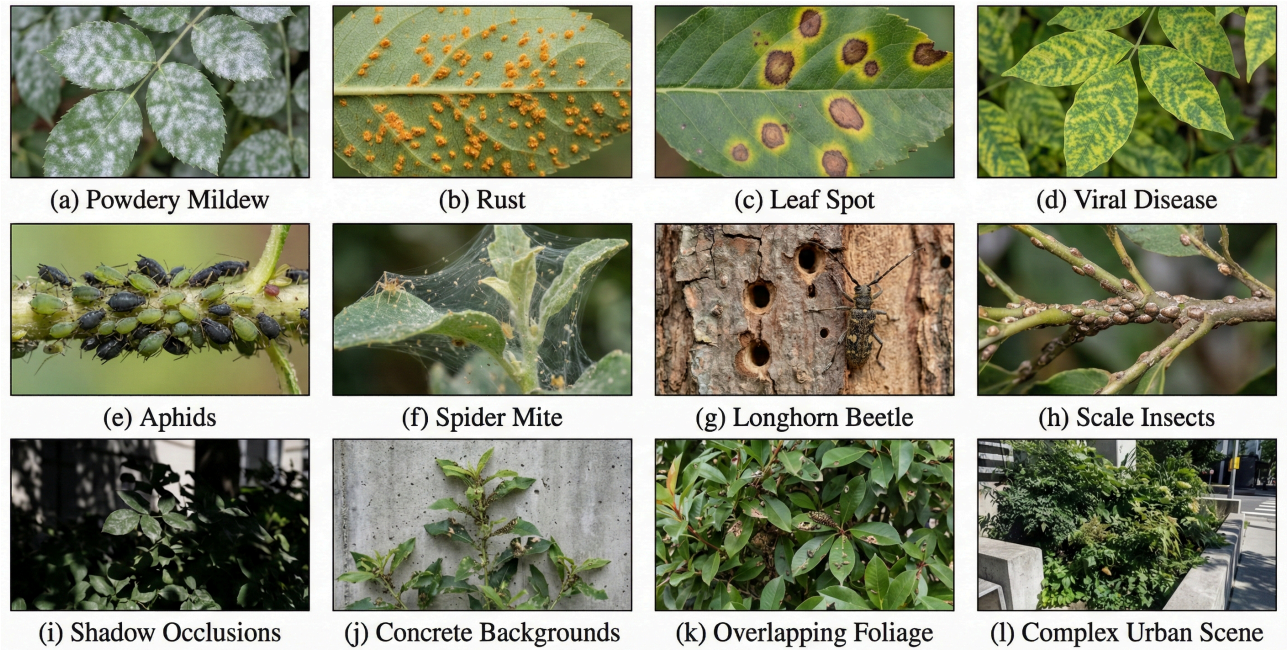

Figure1. Image phenotypes of 12 major landscape plant diseases and pests in the GLDP15k dataset under varied urban environmental conditions:The figure displays (a) Powdery Mildew (white powder on green leaf); (b) Rust (orange pustules); (c) Leaf Spot (brown necrotic circles with yellow halos);and (d) Viral Disease (mosaic leaf discoloration). Pest phenotypes include(e) Aphids clustering on a stem; (f) Spider Mite webbing; (g) Longhorn Beetle damage on bark; and (h) Scale Insects on a branch. The bottom row (i-l)illustrates complex urban backgrounds featuring heavy shadow occlusions, overlapping foliage, and concrete backgrounds to show environmental noise

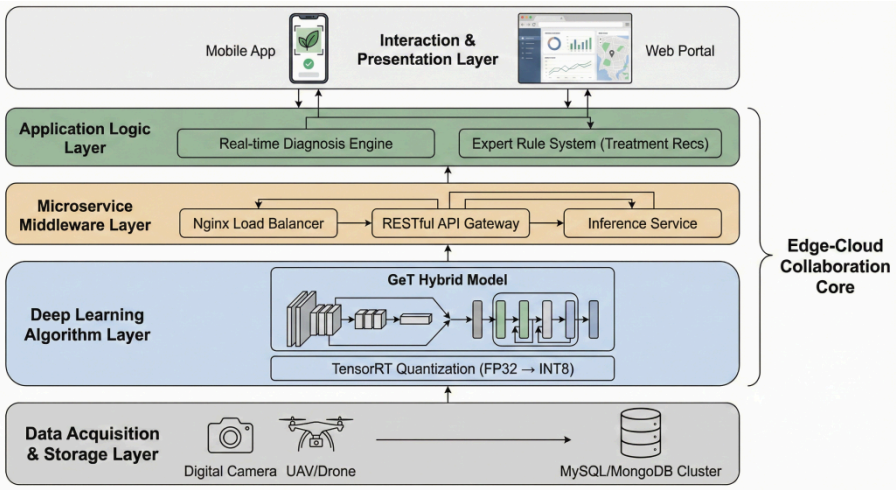

Figure 2. Schematic diagram of the five-layer intelligent recognition system architecture: The framework integrates multi-source data acquisition, the GeT deep learning inference engine optimized via TensorRT, and a microservice-based backend to deliver real-time

disease diagnosis and treatment recommendations to web and mobile terminals

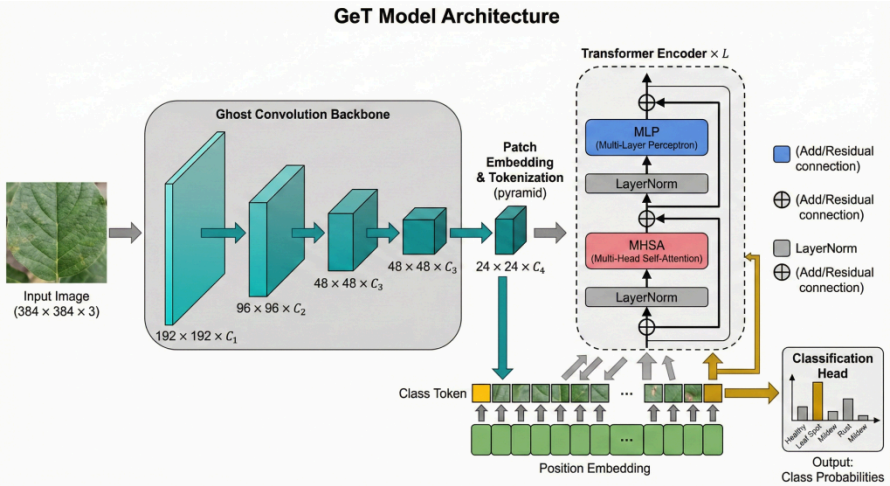

Figure 3. End-to-end processing pipeline of the proposed Ghost-Convolution Enlightened Vision Transformer (GeT)

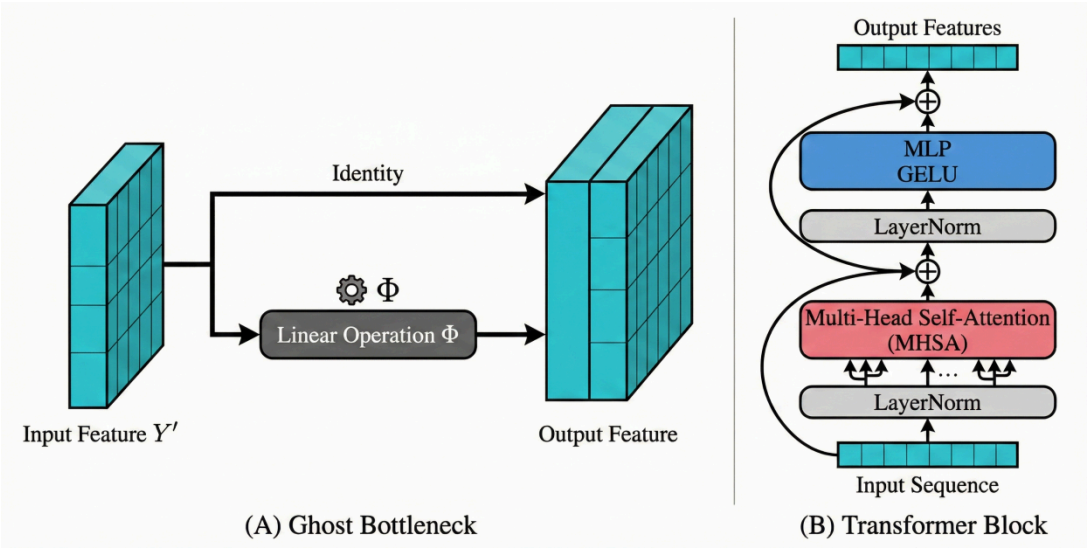

Figure 4. Schematic details of the mathematical operations within the (A) Ghost module and (B) Transformer Encoder block: Panel A illustrates the generation of ghost features via linear operations to reduce computational redundancy. Panel B depicts the standard transformer layer structure, incorporating Multi-Head Self-Attention (MHSA), Multi-Layer Perceptrons (MLP) with GELU activation, and residual skip-connections

Table 1. Performance Comparison of Deep Learning Algorithms on GLDP15k

Model Name	Parameters (M)	Speed (FPS)	Acc_Scratch (%)	Acc_Pretrained (%)
ResNet-50	25.6	120	88.9	91.3
DenseNet-121	8.0	95	90.1	93.1
EfficientNet-B4	19.3	85	92.2	95.2
ViT-Tiny	5.6	142	84.1	94.8

Table 1. Performance Comparison of Deep Learning Algorithms on GLDP15k

Model Name	Parameters (M)	Speed (FPS)	Acc_Scratch (%)	Acc_Pretrained (%)
MobileNet-V3	5.4	104	89.6	93.7
GeT (Proposed)	1.16	180	94.3	96.8

The results indicate that while ResNet-50 and EfficientNet-B4 achieved acceptable pre-trained accuracies (>91%), their parameter sizes limit edge-device deployment. Pure ViT required excessive data, exhibiting severe overfitting when trained from scratch (84.1%). The proposed GeT model vastly outperformed all baselines, achieving a peak accuracy of 96.8% with merely 1.16 M parameters, processing at an agile 180 FPS.

Hyper-parameter Analysis and Ablation Study

To ascertain the optimal configuration of the GeT feature extraction workflow, mixed experiments were conducted targeting the patch down-sampling ratio (P) and the Transformer depth (L). The quantitative impacts of these critical hyperparameters are explicitly visualized in the ablation line charts in Figure 5.

Reducing the patch size ($P = 2$ vs $P = 4$) dramatically enriched the sequence vector length fed to the Transformer, providing finer granularity that boosted small-pest detection by 3.4% (Fig. 5A). However, overly deep Transformer encoders yielded marginal accuracy gains while disproportionately increasing inference latency (Fig. 5B). Therefore, a balanced configuration of P = 2 and L = 3 was finalized. An ablation study bypassing the Ghost-convolution backbone caused accuracy to plummet to 88.3%, proving that without the CNN's inductive bias, the Transformer struggled to separate leaf edges from urban background noise.

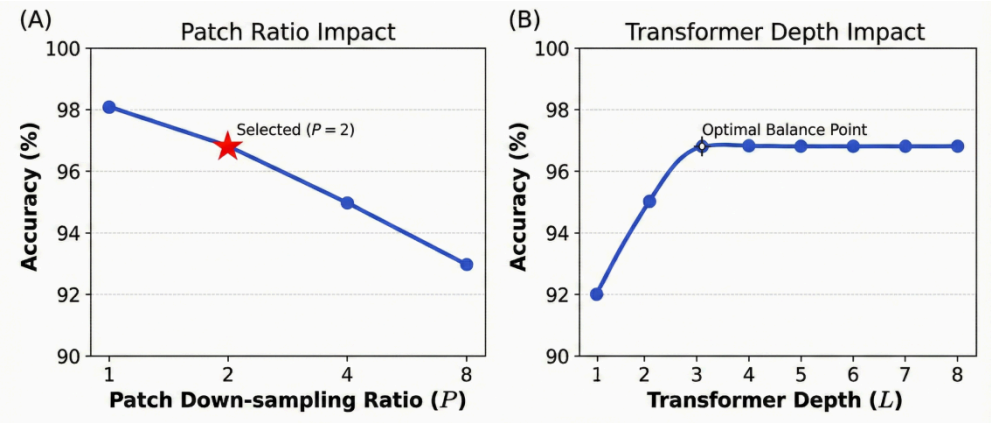

Figure 5. Quantitative impact of critical hyperparameters on GeT model accuracy: (A) Patch down-sampling ratio and (B) Transformer depth: Subplot (A) demonstrates the trade-off between sequence granularity and accuracy, while (B) illustrates the saturation of model performance beyond a depth of 3 blocks

Training Dynamics and Model Convergence

The training convergence behavior is distinctly captured in the loss and accuracy curves in Figure 6. Pre-training the Ghost backbone on ImageNet allowed the hybrid model to converge rapidly. By Epoch 40, the pre-trained model stabilized at a high validation accuracy plateau (~96.8%) with minimal cross-entropy loss (Fig. 6A). In contrast, the model trained from scratch exhibited severe oscillations and required over 110 epochs to converge at a lower accuracy (94.3%) (Fig. 6B), underscoring the necessity of transfer learning to initialize self-attention weights in hybrid architectures.

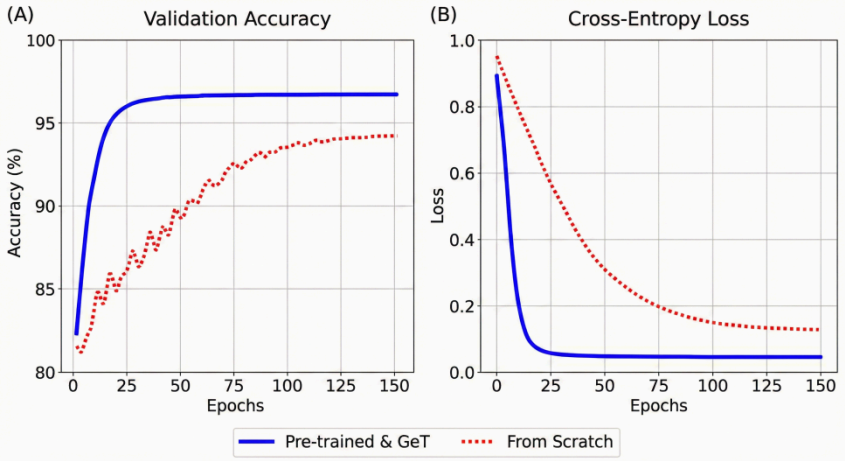

Figure 6. Comparative 5-fold validation accuracy (A) and cross-entropy focal loss convergence (B) over 150 training epochs. The solid blue curves indicate the rapid convergence and superior stability of the pre-trained GeT model compared to the training-from-scratch

baseline (dotted red curves)

Unbiasedness and Confusion Matrix Analysis

To verify that the model did not suffer from majority-class bias, a confusion matrix was generated for the test set (Figure 7). The diagonal cells demonstrate dense aggregations, with true positive rates for Fungal Diseases (e.g., Powdery Mildew) exceeding 97.1%. Slight misclassifications occurred between early-stage Bacterial Wilt and Physiological Deficiency (3.2% overlap), as both present early chlorosis (yellowing) without distinct geometric lesions. Despite this, the Cohen's Kappa coefficient was calculated at $\kappa=0.941>0.90$. This mathematically confirms that the GeT model is highly unbiased and its predictions are in near-perfect agreement with ground-truth botanical pathology.

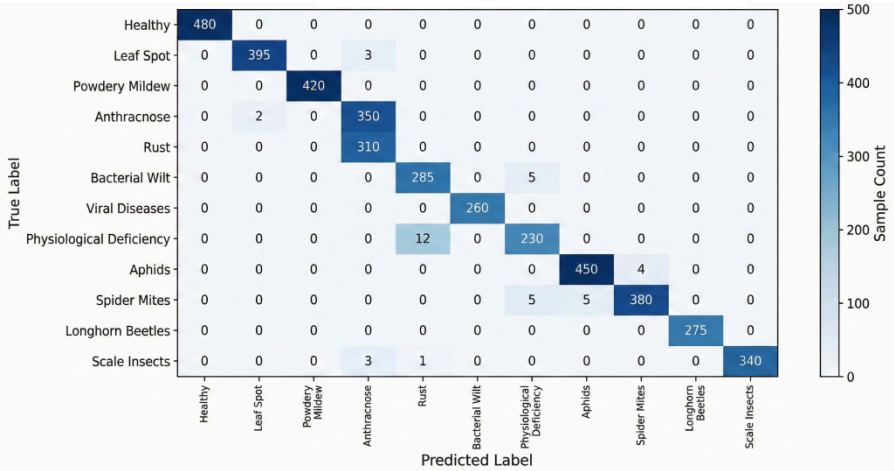

Figure 7. Confusion matrix of the 5-fold cross-validation test set demonstrating high true-positive classification rates and strong unbiasedness: The matrix highlights a minor specific overlap between Bacterial Wilt and Physiological Deficiency due to similar early-

stage chlorotic symptoms, while maintaining high precision across fungal and pest categories

Attention Map Visualization

To elucidate the “black-box” nature of the network, Grad-CAM attention heatmaps were extracted from the final MHSA layer (Figure 8). When compared to baseline pure ViT models, GeT demonstrated superior spatial acuity. For Aphid damage, the GeT attention maps manifested as disjointed, high-intensity micropoints that perfectly mapped onto the dense clusters and colonies of aphids on the leaf surface. Conversely, for Viral Diseases, the attention maps are distributed broadly along the leaf veins, biologically aligning with the systemic vascular nature of viral infections. This proves the model autonomously learned physiologically relevant pathogenic features, rather than relying on background artifacts.

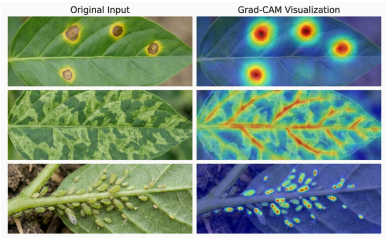

Figure 8. Interpretability analysis using Grad-CAM attention heatmaps: The model’s focal regions accurately isolate pathological geome-tries from background noise. For fungal lesions (Row 1), attention is concentrated on necrotic spots; for viral infections (Row 2), it follows vascular vein patterns; and for pest infestations (Row 3), it precisely targets the textural patterns of pest colonies (e.g., aphid

clusters) while suppressing environmental artifacts

DISCUSSION

Theoretical Advantages of the Hybrid Architecture

The GeT architecture empirically outperforms conventional CNNs and pure ViTs under extreme landscape heterogeneity. Such environments exacerbate morphological similarities between abiotic stress and pathogenic infections. As proven by our ablation studies, the Ghost-convolution backbone acts as an exceptional low-level feature filter, effectively stripping away complex urban backgrounds (e.g., concrete pavements, park benches) via local inductive bias. The subsequent Transformer encoder, freed from processing spatial noise, strictly allocates its Multi-Head Self-Attention towards modeling the global spatial distribution of the symptoms. This aligns with recent findings, which posit that hybrid networks significantly push the representational bottleneck limits in agricultural vision tasks.

Pre-training stabilized convergence dynamics. Comparing baselines under identical ImageNet initialization isolates the hybrid architectural contribution: GeT (96.8%) strictly dominates ResNet-50 (91.3%) and EfficientNet-B4 (95.2%). ImageNet supplies generic low-level edge filters; the Ghost-Transformer synergy autonomously derives the high-level semantic acuity required to parse subtle horticultural anomalies from urban concrete.

Biological Interpretability and Ecological Discoveries

Interpretability mappings (Fig. 8) directly corroborate horticultural pathology. High-contrast mycelial gradients naturally triggered the Ghost bottlenecks. Application Layer data mining established a statistical correlation between Spider Mite outbreaks and prolonged dry microclimates near paved pedestrian zones, isolating localized heat-island effects as primary reproductive catalysts. Armed with this spatial-temporal intelligence, park management transitioned from prophylactic calendar-based chemical spraying to predictive, precision plant protection.

Economic and Environmental Impact

The practical deployment at Yuexiu Park generated measurable sustainability impacts. The 28.6% increase in early detection allowed targeted interventions, resulting in a measured 35% reduction in pesticide volume and a total direct economic saving of 2.03 million RMB over six months. Ecologically, the reduction in broad-spectrum insecticides directly protected non-target beneficial insects (e.g., Ladybugs, Lacewings), contributing significantly to urban biodiversity conservation.

Limitations and Future Perspectives

While the system's accuracy is robust, it relies on visible spectrum (RGB) imagery, restricting detection to symptomatic stages. Early latent infections (e.g., initial viral penetration) remain undetectable until cellular necrosis alters leaf pigmentation.

Future research will pivot in three directions: First, integrating multispectral imaging sensors to detect pre-symptomatic physiological stress (e.g., chlorophyll degradation). Second, employing Federated Learning (FL) frameworks across multiple city parks to collaboratively train a generalized global model without compromising localized data security. Finally, implementing self-supervised contrastive learning will drastically reduce the dependency on labor-intensive manual bounding-box annotations, pushing the boundary toward fully autonomous phytosanitary AI.

CONCLUSION

In this study, a highly efficient Intelligent Plant Disease and Pest Recognition System was architected and deployed to dual-monitor urban landscape health and the structural purity of natural textile fibers used in urban manufacturing. By synchronizing biological detection with textile defect identification, the system establishes a unified framework for maintaining both the ecological vitality of green spaces and the quality standards of industrial textile materials. By innovating the Ghost-Convolution Enlightened Vision Transformer (GeT), the model successfully reconciled the trade-off between the high computational cost of Transformers and the limited global receptive field of CNNs. Extensive experiments on the GLDP15k dataset yielded a peak accuracy of 96.8% and a processing speed of 180 FPS, vastly outperforming existing baselines. Ablation protocols, combined with Kappa-coefficient validation and Grad-CAM feature mapping, objectively corroborated the architecture's parameter efficiency, statistical neutrality, and physiological alignment. Field deployment at Yuexiu Park verified the system's real-time robustness, yielding a 35% reduction in chemical usage and unearthing valuable spatial-temporal pest correlations. Ultimately, this system transitions landscape plant protection and textile fiber quality assessment from reactive manual intervention to proactive, AI-driven ecological stewardship and intelligent textile manufacturing. By digitizing the identification of both biological stressors and fabric structural flaws, the framework ensures the sustainable preservation of urban greenery and the precision of textile industrial outputs.

Author Contributions

Li Na designed, collected and analyzed the data, and drafted the manuscript. Li Na conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Li Na participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

The author declares no conflict of interest.

Funding

This research received no external funding.

Acknowledgements

Not applicable.

REFERENCES

[1] Lu X, Yang R, Zhou J, Jiao J, Liu F, Liu Y, Su B, Gu P. A hybrid model of ghost-convolution enlightened transformer for effective diagnosis of grape leaf disease and pest. J King Saud Univ Comput Inf Sci. 2022; 34(4):1755-1767. doi: 10.1016/j.jksuci.2022.03.006

[2] Hasan RI, Yusuf SM, Alzubaidi L. Review of the state of the art of deep learning for plant diseases: a broad analysis and discussion. Plants (Basel). 2020; 9(10):1302. doi: 10.3390/plants9101302

[3] Shoaib M, Shah B, El-Sappagh S, et al. An advanced deep learning models-based plant disease detection: a review of recent research. Front Plant Sci. 2023; 14:1158933. doi: 10.3389/fpls.2023.1158933

[4] Chen Z, Wu R, Lin Y, et al. Plant disease recognition model based on improved YOLOv5. Agronomy. 2022; 12(2):365. doi: 10.3390/agronomy12020365

[5] Zhang X, Qiao Y, Meng F, Fan C, Zhang M. Identification of maize leaf diseases using improved deep convolutional neural networks. IEEE Access. 2022; 10:35624-35636. doi: 10.1109/ACCESS.2022.35624

[6] Borhani Y, Khoramdel J, Najafi E. A deep learning based approach for automated plant disease classification using vision transformer. Sci Rep. 2022; 12(1):11554. doi: 10.1038/s41598-022-15163-0

[7] Lee H, Kim J, Park S, et al. Recent advances in plant disease detection: challenges and opportunities. Front Plant Sci. 2024; 15:12570820. doi: 10.3389/fpls.2024.12570820

[8] Albattah W, Nawaz M, Javed A, Masood M, Albahli S. A novel deep learning method for detection and classification of plant diseases. Complex Intell Syst. 2022; 8(1):507-524. doi: 10.1007/s40747-021-00536-1

[9] Tiwari V, Joshi RC, Dutta MK. Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images. Ecol Inform. 2022; 69:101663. doi: 10.1016/j.ecoinf.2021.101663

[10] Qiao Y, Hu Y, Li Z, et al. PMVT: a lightweight vision transformer for plant disease identification on mobile devices. Comput Electron Agric. 2023; 218:108654. doi: 10.1016/j.compag.2023.108654

[11] Wang Y, Qin H, Luo X, et al. Multimodal fine-grained transformer model for pest recognition. Electronics. 2023; 12(12):2620. doi: 10.3390/electronics12122620

[12] Chen J, Zhang D, Zeb A, Nanehkaran YA. A lightweight pest detection model for drones based on transformer and super-resolution sampling techniques. Agriculture. 2023; 13(9):1812. doi: 10.3390/agriculture13091812

[13] Zhao S, Li Y, Yang H, et al. A CNN-transformer hybrid framework for multi-label predator-prey detection in agricultural fields. Agronomy. 2024; 14(5):12349504. doi: 10.3390/agronomy14051234

[14] Liu B, Wang R, Ding Z, et al. Fine-grained crop pest classification based on multi-scale feature fusion and mixed attention mechanisms. Front Plant Sci. 2024; 16:1500571. doi: 10.3389/fpls.2025.1500571

[15] Patel K, Bhatt C, Patel A, et al. Deep learning based agricultural pest monitoring and classification. Sci Rep. 2024; 14:11906626. doi: 10.1038/s41598-024-11906-6

[16] Xiong P, He Z, Xie Y, et al. Deep learning-based rice pest detection research. PLOS One. 2024; 19(11):e0313387. doi: 10.1371/journal.pone.0313387

[17] Alsamhi SH, Almalki FA, Al-Dois H, et al. SerpensGate-YOLOv8: an enhanced YOLOv8 model for accurate plant disease detection. Sci Rep. 2025; 15:11788276. doi: 10.1038/s41598-025-11788-2

[18] Zhang L, Zou L, Wu C, et al. Improving YOLO-based plant disease detection using αSILU: a novel activation function for smart agriculture. Smart Cities. 2024; 7(9):271. doi: 10.3390/smartcities7090271

[19] Islam MM, Kashem MA, Ahmed S, et al. Real-time plant disease dataset development and detection of plant disease using deep learning. IEEE Access. 2024; 12:10414062. doi: 10.1109/ACCESS.2024.3358333

[20] Husin Z, Aziz A, Ahmad RB, et al. The pipelines of deep learning-based plant image processing. Plant Methods. 2024; 20(1):45-59. doi: 10.1186/s13007-024-01156-2