RESEARCH ARTICLE

Comparative Study of Lightweight Deep Learning Models for Greenhouse Tomato Leaf Disease Classification Using the Public TLID Dataset

Lihao Ni[1], Fuyin Ye[1], Xiaojun Cui[1], Xiaoman Peng[1], Shaoshuai Song[1] and Wei Luo[1,*]

[1] Wenzhou Academy of Agricultural Sciences, Wenzhou 325006, China

Abstract

Accurate recognition of greenhouse tomato leaf diseases is crucial for crop monitoring, timely intervention, and yield protection. In greenhouse environments, disease symptoms are often affected by complex illumination, background clutter, overlapping leaves, mixed patterns, and subtle inter-class differences, making reliable image-based diagnosis challenging. To evaluate compact convolutional neural networks for this task, this study presents a controlled comparison of five CNN models—MobileNetV3-Large, ShuffleNetV2_x1_0, MobileNetV2, EfficientNet-B0, and ResNet18—using the public Tomato Leaf Image Dataset (TLID). A curated split of 15,254 images covering seven conditions was used, with 10,674 for training, 2,286 for validation, and 2,294 for testing. All models were trained from scratch under identical preprocessing, augmentation, optimization, and selection protocols. Performance was assessed using accuracy, macro-precision/recall/F1, class-wise metrics, confusion matrix, and Grad-CAM visualization. Results show ResNet18 achieved the best overall performance with 71.80% test accuracy and macro-F1 of 0.6726. Among lightweight models, EfficientNet-B0 delivered the strongest results, reaching 70.40% accuracy and macro-F1 of 0.6299, establishing it as the most competitive lightweight baseline. Class-wise analysis indicated Healthy was recognized most reliably, while WhiteFly remained the most challenging due to limited samples, subtle cues, and overlap with healthy leaves. Grad-CAM visualization confirmed the best model focused on symptom-relevant regions rather than background. Overall, findings provide a benchmark for TLID-based classification, identifying EfficientNet-B0 as the strongest lightweight baseline and ResNet18 as the top-performing reference model.

Keywords: greenhouse tomato leaf disease classification, lightweight convolutional neural network, TLID, EfficientNet-B0, ResNet18, Grad-CAM.

1 Introduction

Tomato is one of the most important horticultural crops cultivated in greenhouse systems worldwide, and its productivity is highly vulnerable to foliar diseases

Submitted: 13 February 2026
Accepted: 17 March 2026
Published: 21 March 2026

Vol. 2, **No.** 1, 2026.
10.62762/DIA.2026.103152

***Corresponding author:**
✉ Wei Luo
luowei@radi.ac.cn

Citation
Ni, L., Ye, F., Cui, X., Peng, X., Song, S., & Luo, W. (2026). Comparative Study of Lightweight Deep Learning Models for Greenhouse Tomato Leaf Disease Classification Using the Public TLID Dataset. *Digital Intelligence in Agriculture*, 2(1), 45–53.

and pest-related damage. In practical greenhouse management, delayed recognition of abnormal leaf conditions may lead to disease spread, increased pesticide use, reduced fruit quality, and substantial yield loss. For this reason, early and reliable diagnosis of tomato leaf conditions is of considerable importance for precision agriculture, plant protection, and sustainable cultivation. However, manual inspection remains labor-intensive, observer-dependent, and difficult to scale across large cultivation areas. Its effectiveness may also vary with the experience of field personnel, symptom visibility, and environmental conditions. These limitations have motivated the development of image-based automatic diagnostic methods for crop disease recognition.

With the rapid development of deep learning, convolutional neural networks (CNNs) have significantly advanced plant disease recognition by learning discriminative visual features directly from raw images rather than relying on handcrafted descriptors [1, 2]. Compared with traditional machine-learning pipelines, CNN-based methods often provide stronger robustness and better end-to-end performance, especially when dealing with complex visual patterns and multi-class classification problems. As a result, image-based plant disease analysis has become an active topic in intelligent agriculture and computer vision for agricultural applications.

Nevertheless, high predictive accuracy alone is not sufficient for practical greenhouse deployment. Realistic applications also require models that are computationally efficient, reproducible, and potentially adaptable to resource-constrained hardware platforms. This consideration is particularly relevant in greenhouse monitoring scenarios, where future deployment may involve embedded systems, edge devices, mobile terminals, or lightweight decision-support tools. Accordingly, lightweight CNN architectures are attractive because they may provide a useful trade-off between accuracy and computational complexity. At the same time, it remains necessary to examine whether compact models can maintain sufficient discriminative power when the task involves subtle symptoms, mixed disease categories, and visually similar classes.

This issue is especially important for the Tomato Leaf Image Dataset (TLID), a public dataset collected in greenhouse conditions that includes healthy leaves, disease symptoms, pest-related damage, and mixed categories [3, 4]. Compared with highly controlled laboratory-style datasets, TLID presents a more realistic classification scenario and therefore provides a meaningful test bed for evaluating practical model behavior. In this context, a systematic comparison of representative lightweight and baseline CNN architectures can offer useful evidence for selecting suitable models for greenhouse tomato disease recognition.

Therefore, this study conducts a controlled comparative evaluation of five CNN architectures, namely MobileNetV3-Large, ShuffleNetV2_x1_0, MobileNetV2, EfficientNet-B0, and ResNet18, on the TLID dataset. All models are trained from scratch under a unified experimental protocol so that architectural differences can be assessed more fairly. The main objective is not to propose a new model, but to provide an application-oriented benchmark that compares predictive performance, complexity, class-wise behavior, and qualitative interpretability. The contributions of this work can be summarized as follows. First, it provides a consistent comparative study of five representative CNN architectures on a realistic greenhouse tomato leaf dataset. Second, it analyzes the trade-off between accuracy and computational cost through parameters, FLOPs, and inference-time statistics. Third, it supplements quantitative evaluation with confusion-matrix analysis and Grad-CAM visualization to provide interpretable insights into model behavior.

Plant disease detection using image classification has become an important research direction in intelligent agriculture. Early studies demonstrated that deep learning models can achieve strong disease classification performance on crop image datasets and often outperform traditional handcrafted-feature pipelines [1, 5]. These works helped establish the feasibility of end-to-end image-based disease diagnosis and provided an important foundation for subsequent agricultural vision research.

Subsequent studies further examined tomato leaf disease recognition using pre-trained or customized CNN architectures, confirming the practical value of deep learning for agricultural diagnosis [6]. Such studies showed that CNN-based methods are capable of capturing discriminative symptom patterns from leaf images, but they also highlighted the challenges posed by class similarity, symptom overlap, and differences in real-world image acquisition conditions. In many agricultural scenarios, model performance

is not determined solely by classification capacity, but also by the ability to generalize across variable backgrounds, leaf poses, illumination conditions, and mixed symptom patterns.

More recent research in agricultural vision has also emphasized the importance of compact and deployment-oriented deep models. Lightweight architectures can be beneficial when practical applications require lower computational cost, reduced parameter count, and more convenient deployment on resource-limited systems [7]. However, lightweight design may also reduce representational capacity, which can become problematic when disease symptoms are weak, visually ambiguous, or confounded with pest-related damage. Therefore, comparative evaluation is necessary to determine which compact architectures provide the best balance between predictive performance and model efficiency.

In this context, the TLID/PTLID study introduced a greenhouse tomato dataset and reported promising machine-learning and CNN-based results in a more realistic cultivation setting [3, 4]. Because TLID includes not only healthy and diseased leaves but also pest-related and mixed-condition categories, it provides a more challenging benchmark than simpler disease classification datasets. This characteristic makes TLID particularly suitable for evaluating whether lightweight CNN models can remain competitive under practical greenhouse conditions.

Despite these advances, there remains limited comparative evidence on how widely used lightweight CNNs perform under the same from-scratch training protocol on TLID. Existing work often differs in terms of preprocessing, training settings, model initialization, and evaluation focus, which makes direct comparison difficult. The present study addresses this gap by comparing five representative CNN architectures within a unified experimental framework and by jointly considering predictive performance, model complexity, and qualitative interpretability.

2 Materials and Methods

2.1 Dataset Description

This study used the public Tomato Leaf Image Dataset (TLID), which was released for greenhouse tomato disease and pest analysis and is publicly available through Mendeley Data [4]. The accompanying greenhouse study by Zimmermann et al. [3] described the collection context, acquisition conditions, and practical disease scenarios represented in the dataset. In contrast to idealized laboratory image collections, TLID reflects a more realistic greenhouse environment, where disease diagnosis may be complicated by varying backgrounds, natural leaf arrangement, and symptom overlap.

In the present work, the curated experimental split contained 15,254 images across seven classes: 0-Healthy, 1-Miner, 2-BacterialSpot, 3-PowderyMildew, 4-PowderyMildew_Miner, 5-BacterialSpot_Miner, and 6-WhiteFly. The training, validation, and test subsets contained 10,674, 2,286, and 2,294 images, respectively. As shown in Table 1, most classes contained more than 2,000 images in total, whereas WhiteFly included only 872 images. This made WhiteFly the minority class and a notable challenge for balanced classification. From a practical perspective, this imbalance is meaningful because minority and visually ambiguous categories are often the most difficult cases in agricultural monitoring systems.

2.2 Image Preprocessing and Augmentation

All images were resized to 224 × 224 pixels before being fed into the networks. During training, data augmentation was applied through random horizontal flipping, random rotation, and color jittering in order to improve generalization under natural greenhouse variability. These augmentation operations were intended to expose the models to moderate appearance variation while preserving the essential disease-related visual structure of the leaf images. For validation and test data, only resizing and normalization were applied so that performance evaluation reflected the learned model behavior without additional stochastic perturbation. Image normalization followed the standard ImageNet mean and standard deviation. All compared models shared the same preprocessing pipeline so that observed performance differences could be attributed primarily to architectural differences rather than data handling choices.

2.3 Compared Models and Interpretability Setting

Five CNN architectures were comparatively evaluated in this study: MobileNetV3-Large, ShuffleNetV2_x1_0, MobileNetV2, EfficientNet-B0, and ResNet18. MobileNetV2 and MobileNetV3 are widely adopted lightweight networks designed for efficient inference on resource-constrained devices [8, 9]. ShuffleNetV2 was developed according to practical efficiency guidelines and is often regarded as a highly

Table 1. Class distribution of the TLID dataset used in this study.

Class	Total	Train	Validation	Test
Healthy	2688	1881	403	404
Miner	2255	1578	338	339
BacterialSpot	2395	1676	359	360
PowderyMildew	2181	1526	327	328
PowderyMildew_Miner	2361	1652	354	355
BacterialSpot_Miner	2502	1751	375	376
WhiteFly	872	610	130	132
Total	**15254**	**10674**	**2286**	**2294**

compact architecture for low-cost inference [10]. EfficientNet-B0 scales depth, width, and resolution in a balanced manner and has been widely used as a strong compact baseline [11]. ResNet18, although not the lightest model in the present comparison, was included as a reference benchmark because residual learning remains a stable and competitive backbone for image classification tasks [12].

For all models, the original classifier head was replaced with a seven-class output layer corresponding to the TLID label space. This ensured that all architectures were directly comparable under the same classification objective. To supplement quantitative performance analysis with qualitative interpretability, Grad-CAM was applied to the best-performing model, ResNet18, using representative correctly classified and misclassified test samples [13]. This visualization step was intended to provide preliminary evidence regarding whether the model relied on symptom-relevant leaf regions rather than irrelevant background content.

2.4 Experimental Protocol

The experimental pipeline was designed to remain simple, reproducible, and directly comparable across all evaluated models. Each network was trained from scratch under the same data split, augmentation strategy, optimizer configuration, and stopping criterion. Best-model selection was based on validation macro-F1, and the selected checkpoint for each architecture was then used for final evaluation on the independent test set. This design was chosen to reduce confounding factors and to ensure that the reported differences mainly reflected architectural behavior under a unified experimental setting.

The decision to train all models from scratch was made to preserve consistency across the comparison. Although transfer learning from ImageNet pretraining is common in agricultural image classification, pretrained initialization may introduce an additional source of variation when the goal is to compare architectural behavior as fairly as possible. Therefore, the present work focused on a unified from-scratch setting as a controlled benchmark rather than an attempt to maximize absolute performance through architecture-specific tuning.

3 Results

3.1 Training Settings

All experiments were implemented in PyTorch and executed on a server equipped with an NVIDIA TITAN Xp GPU. One GPU was used during training. The Python version was 3.10.20 and the PyTorch version was 2.5.1+cu121. Cross-entropy loss was used for optimization, and AdamW was adopted with a learning rate of 1×10^{-4}. The batch size was set to 48, the maximum number of epochs was 20, and early stopping with a patience value of 5 was employed to reduce overfitting. Randomly initialized model weights were used in the main comparison so that all networks were evaluated under the same from-scratch learning condition.

The overall training configuration was intentionally kept moderate rather than aggressively tuned for a single model. This choice made the comparison more transparent and helped reduce the likelihood that performance differences were driven by architecture-specific hyperparameter advantages. Accordingly, the reported results should be understood as comparative evidence under a unified training regime.

3.2 Evaluation Metrics

Performance was evaluated using test accuracy, macro-precision, macro-recall, macro-F1, class-wise precision, recall, and F1-score, as well as normalized confusion matrices and qualitative interpretability analysis. Among these metrics, macro-F1 was treated

Table 2. Performance and model complexity comparison of different CNN models on the TLID dataset.

Model	Best Val Macro-F1	Test Acc.(%)	Test Macro-F1	Params (M)	FLOPs (G)	Inference Time(ms/image)
MobileNetV3-Large	0.5715	63.78	0.5603	4.2110	0.4307	7.7348
ShuffleNetV2_x1_0	0.5626	64.82	0.5663	1.2608	0.3034	7.3159
MobileNetV2	0.5910	66.43	0.5830	2.2328	0.6524	6.2000
EfficientNet-B0	0.6445	70.40	0.6299	4.0165	0.8277	9.4474
ResNet18	0.6859	71.80	0.6726	11.1801	3.6471	2.6754

as the primary selection criterion because it assigns equal importance to each class and is therefore more informative than accuracy alone for moderately imbalanced datasets such as TLID. In agricultural classification tasks where minority classes may carry substantial practical importance, macro-F1 provides a more balanced summary of model behavior.

In addition to predictive performance, model complexity was measured using the number of parameters, FLOPs, and single-image inference time under the current hardware environment. These indicators were used to assess the trade-off between classification quality and computational cost. It should be noted, however, that measured latency is dependent on the hardware-software stack and low-level implementation details. Therefore, inference time was treated only as a practical runtime reference rather than a definitive indicator of model lightweightness.

3.3 Performance Comparison

The comparative experiments revealed clear performance differences among the five evaluated CNN models. As summarized in Table 2, ResNet18 achieved the best overall performance, with a best validation macro-F1 of 0.6859, a test accuracy of 71.80%, and a test macro-F1 of 0.6726. These results indicate that, under the present from-scratch training setting, ResNet18 provided the strongest overall balance of feature extraction capability and classification reliability. Among the lightweight models, EfficientNet-B0 performed best, reaching a best validation macro-F1 of 0.6445, a test accuracy of 70.40%, and a test macro-F1 of 0.6299. This result suggests that EfficientNet-B0 was the most competitive compact architecture in the present comparison.

MobileNetV2 showed moderate performance, with a test macro-F1 of 0.5830, indicating that it remained a viable lightweight baseline but was clearly less competitive than EfficientNet-B0. ShuffleNetV2_x1_0 and MobileNetV3-Large produced weaker results, with test macro-F1 scores of 0.5663 and 0.5603, respectively. Overall, the ranking pattern indicates that model capacity and representational strength still matter when the task involves visually similar symptoms, mixed disease classes, and minority categories. Although compact architectures can offer considerable efficiency benefits, not all lightweight models are equally effective for challenging greenhouse leaf classification. Class-wise analysis further showed that Healthy, Miner, and BacterialSpot were generally easier to classify, whereas WhiteFly remained the most challenging class across models. For the best-performing ResNet18 model, the F1-scores for Healthy, Miner, BacterialSpot, PowderyMildew, PowderyMildew_Miner, BacterialSpot_Miner, and WhiteFly were 0.86, 0.73, 0.73, 0.69, 0.77, 0.63, and 0.31, respectively. This result indicates a large gap in difficulty across categories and highlights the practical importance of minority-class robustness. In particular, the low F1-score for WhiteFly suggests that limited sample size and weak symptom distinctiveness can substantially impair classification performance, even for the strongest model in the comparison.

Figure 1. Normalized confusion matrix of ResNet18 on the TLID test set.

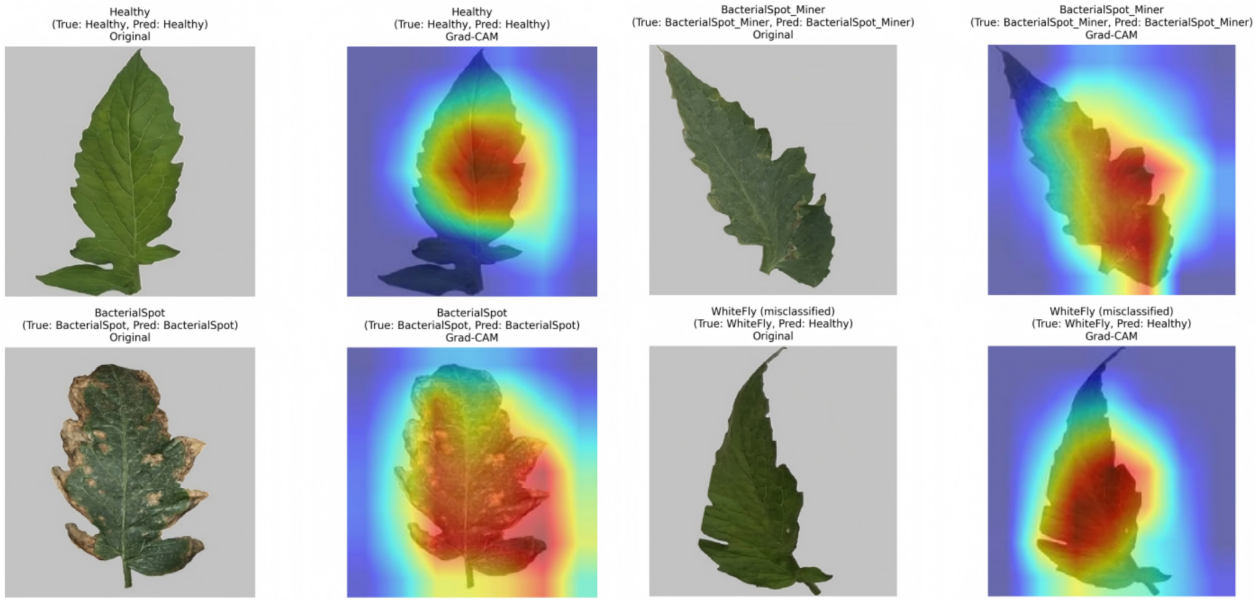

Figure 2. Grad-CAM visualizations of ResNet18 on representative TLID test samples, including correctly classified Healthy, BacterialSpot, and BacterialSpot_Miner leaves, as well as a WhiteFly sample misclassified as Healthy.

The normalized confusion matrix of ResNet18, shown in Figure 1, provides additional insight into misclassification behavior. The major confusion patterns were primarily observed between BacterialSpot and BacterialSpot_Miner, as well as between PowderyMildew and related classes, whereas WhiteFly remained the most challenging class. These confusion patterns suggest that mixed-symptom categories and visually overlapping disease conditions remain difficult even when overall model performance is relatively strong. In practical greenhouse monitoring, such errors are important because mixed or subtle symptoms may require more cautious diagnostic support rather than direct reliance on a single predicted label.

Table 2 also summarizes the joint comparison of predictive performance and model complexity. ResNet18 produced the best predictive results but also had the largest parameter count and FLOPs, indicating that its superior performance came at a higher representational cost. ShuffleNetV2_x1_0 was the most compact model in terms of parameters and FLOPs, whereas MobileNetV2 provided a somewhat stronger but still lightweight baseline. Among the lightweight models, EfficientNet-B0 offered the best trade-off between classification quality and model complexity. Although its measured inference time was longer than that of some other models under the present hardware environment, its overall predictive advantage makes it the most competitive lightweight baseline in this study. The measured inference time should therefore be interpreted only as a practical runtime reference under the current hardware and software environment rather than as a sole criterion for model selection.

Grad-CAM visualization further supported the quantitative findings. As shown in Figure 2, ResNet18 generally concentrated on leaf regions and symptom-relevant areas in correctly classified samples, including representative Healthy, BacterialSpot, and BacterialSpot_Miner leaves. By contrast, the misclassified WhiteFly case showed a weaker and less discriminative attention pattern. This observation suggests that the model was able to identify meaningful disease-related regions when the visual signal was relatively clear, but had greater difficulty when symptom evidence was weak, ambiguous, or visually similar to healthy tissue. Therefore, the interpretability results are broadly consistent with the confusion-matrix analysis and class-wise performance patterns.

4 Discussion

The present results should be interpreted in the context of prior tomato leaf disease studies. A number of previous studies reported very high classification performance, but most of them were conducted on PlantVillage-type or otherwise relatively controlled datasets, often with transfer learning or architecture-specific optimization. For example, Tan et al. [6] compared classical machine learning and deep learning methods on PlantVillage tomato

images and reported that ResNet34 achieved 99.7% accuracy, EfficientNet-B0 98.9%, and MobileNetV2 91.2%. Likewise, Attallah [14] proposed a compact CNN framework with transfer learning, feature fusion, and feature selection and reported up to 99.92% accuracy. Other tomato-specific studies also reached similarly high performance by incorporating attention mechanisms, pruning, or customized lightweight designs, such as the lightweight attention-based CNN of Bhujel et al. [15] and the LBFNet model of Chen et al. [16]. In addition, Saeed et al. [17] reported up to 99.22% accuracy using pretrained Inception-family models on a dataset that combined PlantVillage and field-recorded images. These studies clearly demonstrate the strong potential of deep learning for tomato disease recognition.

However, such results are not directly comparable to the present study. Our work was conducted on the public TLID dataset, which was collected in greenhouse conditions and includes not only healthy and diseased leaves, but also pest-related and mixed-condition categories. The original TLID/PTLID study by Zimmermann et al. [3] already showed that realistic greenhouse tomato images constitute a more challenging problem than conventional laboratory-style datasets, and their best reported result reached 90.48% accuracy when patch-based PTLID images and voting-based combinations of CNN models were used. By contrast, the present work focused on whole-leaf TLID classification and trained all five models from scratch under the same preprocessing, augmentation, optimization, and model-selection protocol. Therefore, our lower absolute performance should be interpreted in light of a more difficult dataset setting and a stricter controlled-comparison design rather than as evidence of inferior model behavior per se.

From this perspective, the main strength of the present study is not the proposal of a new state-of-the-art architecture, but the provision of a fair and reproducible benchmark for lightweight CNN comparison under realistic greenhouse conditions. Within this unified from-scratch setting, ResNet18 achieved the best overall performance, whereas EfficientNet-B0 emerged as the strongest lightweight baseline. This finding is consistent with previous literature showing that EfficientNet-style scaling can offer a favorable balance between efficiency and predictive performance [6, 14], while also indicating that additional representational capacity remains important when the task involves subtle visual cues, mixed symptoms, and minority classes. In other words, our study contributes comparative evidence that is especially relevant for deployment-oriented model selection on TLID-like greenhouse datasets, where direct transfer of conclusions from PlantVillage-based studies may be unreliable.

The class-wise results further clarify why this task remains difficult. WhiteFly was consistently the most challenging class, which is likely related to its smaller sample size and weaker visual distinctiveness. Similar observations have been made in prior real-world or complex-environment studies, where inter-class similarity, background interference, and symptom ambiguity remain major performance bottlenecks [3, 15–17]. In our case, the confusion between BacterialSpot and BacterialSpot_Miner, as well as between PowderyMildew and related classes, suggests that mixed-symptom categories deserve more focused treatment in future work. Possible directions include class-balanced optimization, targeted augmentation for minority categories, lesion-aware or region-guided learning, and hierarchical classification strategies.

Another important implication concerns the interpretation of efficiency. Some previous studies improved portability by using lightweight attention modules, pruning, or compact custom designs [15, 16], whereas others pursued stronger absolute accuracy through transfer learning or model fusion [14, 17]. In our experiments, EfficientNet-B0 offered the best overall trade-off among the lightweight backbones, but the latency results should not be overinterpreted, because runtime is highly dependent on the specific hardware-software environment. For this reason, parameters and FLOPs remain more appropriate structural indicators of lightweightness, while the reported inference time should be regarded only as a runtime reference under the present experimental platform.

Overall, the present study complements prior tomato disease literature by filling a different niche. Rather than maximizing accuracy on a controlled or heavily optimized setting, it establishes a controlled benchmark on a recent greenhouse tomato dataset under a unified from-scratch protocol. This makes the results particularly useful for researchers who need a realistic reference point for TLID-based model selection and for future studies that aim to improve minority-class robustness, mixed-symptom discrimination, and deployment readiness in greenhouse environments.

5 Conclusion

This study comparatively evaluated five CNN models—MobileNetV3-Large, ShuffleNetV2_x1_0, MobileNetV2, EfficientNet-B0, and ResNet18—for greenhouse tomato leaf disease classification using the public TLID dataset. Under a unified experimental protocol, the comparison revealed clear performance differences among the evaluated architectures and provided a practical benchmark for model selection in this task.

ResNet18 achieved the best overall classification performance, with a test accuracy of 71.80% and a macro-F1 of 0.6726. Among the lightweight models, EfficientNet-B0 showed the strongest and most balanced performance, reaching a test accuracy of 70.40% and a macro-F1 of 0.6299. MobileNetV2 produced moderate results, whereas ShuffleNetV2_x1_0 and MobileNetV3-Large were weaker on this dataset. These findings indicate that lightweight CNNs are feasible for greenhouse tomato leaf disease classification, but that model capacity remains important for distinguishing visually similar, mixed-symptom, and minority categories under realistic greenhouse conditions.

Class-wise analysis and confusion-matrix results further showed that Healthy, Miner, and BacterialSpot were easier to recognize, whereas WhiteFly remained the most challenging class. Grad-CAM visualization provided preliminary interpretability support for the best-performing model by showing that it generally relied on leaf-related and symptom-relevant regions rather than irrelevant background content. Taken together, the results suggest that EfficientNet-B0 can be regarded as the strongest lightweight baseline under the present setting, while ResNet18 serves as the highest-performing reference benchmark for TLID-based greenhouse tomato leaf disease classification. Future work should focus on dataset expansion, minority-class improvement, cross-dataset validation, and more deployment-oriented evaluation on practical greenhouse hardware.

Data Availability Statement

The TLID dataset used in this study is publicly available at https://doi.org/10.17632/kt64b2kh89. The final processed data split, trained model weights, and source code are available at the public GitHub repository: https://github.com/nilihao2003/tomato-leaf-classification.git

Funding

This work was supported in part by the Wenzhou Basic Scientific Research Project under Grant GG20250197, and in part by the Zhejiang Provincial Science and Technology Commissioner Project under Grant 2025CNYJY04.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Mohanty, S. P., Hughes, D. P., & Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419. [CrossRef]

[2] Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*.

[3] Zimmermann, G. B., Pellenz, M. E., Costa, Y. M. G., & Britto Jr, A. S. (2025). Enhancing disease and pest detection in greenhouse tomato cultivation using advanced machine learning on new dataset of images. *Journal of the Brazilian Computer Society, 31*(1), 187-202. [CrossRef]

[4] Pellenz, M., Zimmermann, G. B., Britto Jr, A. S., & Costa, Y. M. G. (2025). Tomato Leaf Image Dataset (TLID/PTLID) [Dataset]. *Mendeley Data, V2*. [CrossRef]

[5] Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep learning for tomato diseases: Classification and symptoms visualization. *Applied Artificial Intelligence, 31*(4), 299-315. [CrossRef]

[6] Tan, L., Lu, J., & Jiang, H. (2021). Tomato leaf diseases classification based on leaf images: A comparison between classical machine learning and deep learning methods. *AgriEngineering, 3*(3), 542-558. [CrossRef]

[7] Zhang, Y., Wu, G., & Shen, J. (2026). Precise tea leaf disease detection using UAV low-altitude remote sensing and optimized YOLO11 model. *PLoS One, 21*(2), e0342545. [CrossRef]

[8] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (pp. 4510-4520). [CrossRef]

[9] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L. C., Tan, M., ... & Le, Q. (2019, October). Searching for MobileNetV3. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1314-1324). IEEE. [CrossRef]

[10] Ma, N., Zhang, X., Zheng, H. T., & Sun, J. (2018, September). ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In *European Conference on Computer Vision* (pp. 122-138). [CrossRef]

[11] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). IEEE. [CrossRef]

[13] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 618-626). [CrossRef]

[14] Attallah, O. (2023). Tomato Leaf Disease Classification via Compact Convolutional Neural Networks with Transfer Learning and Feature Selection. *Horticulturae, 9*(2), 149. [CrossRef]

[15] Bhujel, A., Kim, N. E., Arulmozhi, E., Basak, J. K., & Kim, H. T. (2022). A lightweight attention-based convolutional neural networks for tomato leaf disease classification. *Agriculture, 12*(2), 228. [CrossRef]

[16] Chen, H., Wang, Y., Jiang, P., Zhang, R., & Peng, J. (2023). LBFNet: A Tomato Leaf Disease Identification Model Based on Three-Channel Attention Mechanism and Quantitative Pruning. *Applied Sciences, 13*(9), 5589. [CrossRef]

[17] Saeed, A., Abdel-Aziz, A. A., Mossad, A., Abdelhamid, M. A., Alkhaled, A. Y., & Mayhoub, M. (2023). Smart Detection of Tomato Leaf Diseases Using Transfer Learning-Based Convolutional Neural Networks. *Agriculture, 13*(1), 139. [CrossRef]

Lihao Ni is a Senior Engineer at Wenzhou Academy of Agricultural Sciences, Wenzhou 325006, China. His research interests include blockchain technology, artificial intelligence, smart agriculture, and their applications in vocational education and digital agriculture. (Email: nilihao@wzvcst.edu.cn)

Fuyin Ye is a Lecturer at Wenzhou Academy of Agricultural Sciences, Wenzhou 325006, China. His research interests include deep learning, image processing, smart agriculture, and hyperspectral image classification. (Email: yefuyin@wzvcst.edu.cn)

Xiaojun Cui is a Professor at Wenzhou Academy of Agricultural Sciences, Wenzhou 325006, China. He holds a Ph.D. in Engineering. His research interests include computer-related teaching and research, artificial intelligence, and smart agriculture. (Email: cuixiaojun@wzvcst.edu.cn)

Xiaoman Peng is a Teaching Assistant at Wenzhou Academy of Agricultural Sciences, Wenzhou 325006, China. She holds a Master's degree in Electronic Information. Her research interests include multimodal information retrieval, video temporal grounding, fine-grained image analysis, and smart agriculture. (Email: pengxiaoman@wzvcst.edu.cn)

Shaoshuai Song is a Lecturer at Wenzhou Academy of Agricultural Sciences, Wenzhou 325006, China. His research interests include software technology development, blockchain technology application, and smart agriculture. (Email: songshaoshuai@wzvcst.edu.cn)

Wei Luo received his Ph.D. degree in Cartography and Geographic Information Systems from the University of Chinese Academy of Sciences (UCAS). He is currently an Associate Professor at Wenzhou Academy of Agricultural Sciences, Wenzhou 325006, China. His research interests include artificial intelligence, image processing, cartography and geographic information systems, and smart agriculture. (Email: luowei@radi.ac.cn)