OPEN ACCESS
International Journal of Innovations in Science & Technology

Deep Transfer Learning with Hybrid CNN Fusion for Smart Tomato Leaf Disease Diagnosis

Ahmed Waheed Awan[1, 4], Syed Muhammad Ali[2], Osama Zafar[3], Hania Siddiqui[2], Muhammad Salman[1, 4], Armughan Ali[1, 2,*]

[1]Department of Electrical Engineering, Wah Engineering College, University of Wah, Wah Cantt 47040, Pakistan.

[2]Department of Computer Science, University of Wah, Wah Cantt 47040, Pakistan.

[3]Department of Computer Science, COMSATS University Islamabad, Pakistan.

*Correspondence: armughan.ali@wecuw.edu.pk

Citation | Awan. A. W, Ali. S. M, Zafar. O. Salman. M, Ali. A, "Deep Transfer Learning with Hybrid CNN Fusion for Smart Tomato Leaf Disease Diagnosis", IJIST, Special Issue pp 608-622, May 2026

Received | April 04, 2026 **Revised |** May 12, 2026 **Accepted |** May 14, 2026 **Published |** May 17, 2026.

Tomato plants are highly susceptible to a wide range of bacterial, viral, and fungal diseases that significantly affect crop yield and agricultural productivity, making rapid and accurate identification essential for effective crop management and early intervention. This paper presents Fusion-TLNet, a hybrid deep learning framework that integrates multiple transfer learning models using a feature-level fusion strategy for robust tomato leaf disease classification. The proposed architecture combines three pretrained convolutional neural networks: EfficientNetB0, MobileNetV3-Small, and DenseNet121 to capture complementary hierarchical representations of texture, shape, and color features from tomato leaf images. Experimental results on the PlantVillage dataset demonstrate that Fusion-TLNet achieves an accuracy of 99.34%, precision of 99.28%, recall of 99.22%, and F1-score of 99.25%, outperforming individual backbone models by up to 1.23% in accuracy. In addition, the model maintains a low prediction latency of 9.7 ms per image, making it suitable for real-time deployment in resource-constrained agricultural environments. The proposed model provides an interpretable, efficient, and scalable solution for intelligent plant disease diagnosis, supporting the advancement of data-driven precision agriculture.

Keywords: Tomato Leaf Disease Detection, Deep Learning, Transfer Learning, CNN Fusion, Efficient Net, Mobile Net, Dense Net, Feature-Level Fusion, Precision Agriculture, Explainable AI.

Introduction:

Tomato (Solanum lycopersicum) is one of the most economically important vegetable crops worldwide, contributing significantly to the global food supply and agricultural economies. However, tomato plants are highly susceptible to a wide range of bacterial, viral, and fungal diseases that can severely reduce crop yield and quality [1][2]. Early and accurate disease detection is therefore essential for effective crop management and minimizing economic losses. Traditional manual inspection by agricultural experts is often time-consuming, subjective, and impractical for large-scale monitoring. Recent advances in artificial intelligence (AI), computer vision, and deep learning have enabled automated plant disease detection systems capable of analyzing leaf images with high accuracy and efficiency [3][4], making these systems promising tools for precision agriculture.

Several studies have explored deep learning techniques for tomato leaf disease detection. [5] Proposed a transfer learning–based CNN for identifying multiple tomato diseases from leaf images. [6] Introduced the ToLeD model, which achieved an average accuracy of 91.2% across nine disease classes and one healthy class. [7] Developed a hybrid framework combining pretrained CNN feature extraction with machine learning classifiers, achieving accuracies above 99%. Similarly, [8] integrated convolutional attention modules with an SVM classifier to improve computational efficiency for mobile deployment. These studies demonstrate the effectiveness of deep learning and transfer learning for automated plant disease diagnosis.

Despite these advances, several challenges remain in existing approaches. Many models rely heavily on controlled datasets and may not perform well under real-world conditions involving varying illumination, occlusions, and complex backgrounds. Additionally, high-performing deep learning models often require significant computational resources, limiting their deployment on low-power agricultural devices. Another limitation is the lack of interpretability in many systems, which reduces user trust and practical adoption in farming environments. These challenges highlight the need for a lightweight, robust, and interpretable solution that can achieve high accuracy while maintaining real-time performance and deployment feasibility.

Research Objectives:

The primary objective of this study is to develop an accurate, efficient, and interpretable deep learning framework for automated tomato leaf disease classification. The specific objectives of this research are as follows:

To design a hybrid transfer learning architecture that integrates multiple CNN backbones (EfficientNetB0, MobileNetV3-Small, and DenseNet121) for enhanced feature representation.

To improve classification accuracy by leveraging feature-level fusion of complementary deep features extracted from different pretrained models.

To reduce computational complexity and ensure suitability for real-time deployment on resource-constrained devices such as mobile and edge-based agricultural systems.

To evaluate the robustness and generalization capability of the proposed model under varying environmental conditions.

Novelty of the Study:

The proposed study introduces several key contributions that distinguish it from existing tomato leaf disease classification approaches:

A **hybrid multi-backbone fusion framework** (Fusion-TLNet) that integrates EfficientNetB0, MobileNetV3-Small, and DenseNet121 at the feature level to exploit complementary representations.

Unlike traditional ensemble methods, the proposed approach performs **compact feature-level fusion** instead of decision-level aggregation, resulting in improved accuracy with reduced computational overhead.

The model is specifically designed to achieve a **balance between high accuracy and low inference latency**, making it suitable for real-time deployment on edge devices such as smartphones and IoT-based agricultural systems.

Integration of **Grad-CAM-based interpretability** to enhance transparency and trust by highlighting disease-relevant regions in leaf images.

A comprehensive evaluation including **performance comparison, ablation analysis, efficiency analysis, and robustness testing**, demonstrating the effectiveness and reliability of the proposed framework.

Related Work:

Tomato leaf disease detection has gained significant attention in precision agriculture due to its importance for improving crop productivity and sustainability. Over time, research has evolved from handcrafted feature extraction methods toward deep learning and explainable AI (XAI) frameworks that provide higher accuracy and improved interpretability, as summarized in Table 1.

Early studies primarily relied on handcrafted image features. [1] Employed a hybrid pipeline using discrete wavelet transform (DWT), gray-level co-occurrence matrix (GLCM), and principal component analysis (PCA), achieving 99.97% accuracy but with high computational complexity. With the emergence of deep learning, convolutional neural networks (CNNs) became dominant. [9] Trained a CNN on tomato leaf images covering nine diseases and achieved 98.49% accuracy. Similarly, transfer learning approaches using architectures such as VGG16 and ResNet have been explored by [10] and [11], demonstrating improved classification performance by leveraging pretrained visual representations.

Recent research has focused on lightweight and efficient architectures suitable for real-world deployment. [12] Proposed a MobileNetV2-based model achieving 99.30% accuracy with a compact model size, making it suitable for resource-constrained devices. [13] Introduced DGP-SNNet, which integrates grouped attention and neuron selectivity transfer to improve efficiency while maintaining high accuracy. [14] proposed XLTLDisNet for lightweight disease detection, while [2] presented an SDG-driven CNN framework aligning agricultural AI with sustainability objectives.

Interpretability has also become an important aspect of modern agricultural AI systems. [4] Introduced XSE-TomatoNet, an EfficientNetB0-based architecture integrating squeeze-and-excitation modules with explainability tools such as LIME, SHAP, and Grad-CAM++. Similarly, [15] combined CNN feature extraction with SVM classification to enhance interpretability while maintaining competitive accuracy.

More recently, transformer-based and hybrid architectures have been explored for improved feature representation and classification performance. [3] Applied Vision Transformers (ViT) to tomato disease detection, achieving strong performance on both controlled and field datasets. [16] Further extended this approach by integrating transformer architectures with contextual reasoning. In addition, [17] introduced an attention-based T-LSTM framework that achieved near-perfect classification accuracy.

Furthermore, recent advancements highlight the effectiveness of hybrid deep learning models that combine convolutional neural networks with transformer-based architectures to capture both local and global features. For instance, [18] proposed the X-ViTCNN framework, which integrates DenseNet201 and MobileNetV2 with a customized Vision Transformer for multi-stage Alzheimer's disease prediction. The model effectively fuses local structural features with global contextual representations, resulting in improved classification accuracy and robustness. Additionally, Grad-CAM-based visualization was employed to enhance interpretability and provide insights into the model's decision-making process. This demonstrates the growing importance of hybrid and explainable architectures, which can be effectively extended to agricultural disease detection tasks. Real-time detection systems have also been proposed for practical agricultural deployment. [19] Developed LT-YOLOv10n, a lightweight detector capable of real-time inference on embedded hardware. Similarly, [20] proposed CONF-RCNN for field-based disease detection, while [21] introduced E-Tomato Det using CSW within a Transformer architecture for enhanced global–local feature perception.

Despite these advances, several challenges remain. Many models rely on controlled datasets and may struggle under real-world environmental variations such as illumination changes, occlusion, and complex backgrounds. Additionally, balancing model accuracy with computational efficiency remains critical for deployment on resource-constrained agricultural devices. These limitations motivate the development of lightweight and robust deep learning frameworks capable of maintaining high performance while supporting real-world agricultural applications.

Methodology:

This section describes the architecture, training process, and evaluation pipeline of the proposed Fusion-TLNet model, an optimized transfer learning and feature fusion framework designed for accurate and computationally efficient tomato leaf disease classification. The proposed approach leverages the representational diversity of multiple pre-trained convolutional neural networks (CNNs) to achieve robust performance across heterogeneous imaging conditions.

Overview of the Proposed Framework:

The complete workflow of the proposed system is illustrated in Figure 1 (conceptually consisting of four sequential stages): (1) image acquisition and preprocessing, (2) deep feature extraction through transfer learning backbones, (3) multi-stream feature fusion and dimensionality reduction, and (4) classification through a lightweight dense head.

The proposed framework aims to address three major challenges observed in traditional deep learning models for plant disease classification: (i) overfitting on small datasets, (ii) limited generalization under varying illumination and occlusion, and (iii) lack of computational scalability on low-end agricultural devices.

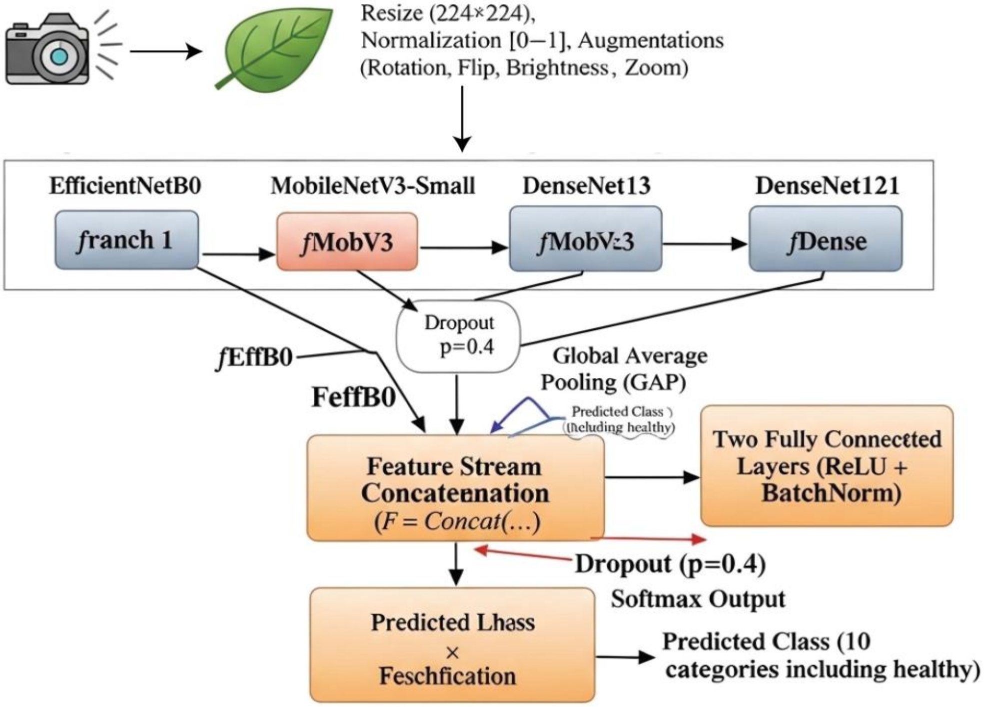


Figure 1. Proposed framework architecture of Fusion-TLNet.

Table 1. Comparative summary of recent tomato leaf disease detection models (reversed order: latest to earliest).

Study	**Model /Technique**	**Dataset**	**Accuracy (%)**	**Key Features**	**Explainability / Deployment**
Chelladurai et al. (2025)	T-LSTM + Attention + U-Net	PlantVillage	99.98	Sequential learning, attention	Highest reported accuracy
Abebe et al. (2025)	SDG-driven CNN	Mixed (local +public)	99.8	Custom CNN + VGG16 hybrid	Sustainability oriented
Bellout et al. (2025)	LT-YOLOv10n (CBAM + C3f)	Field dataset	98.7 (mAP_{50})	IoT-enabled detection	Real-time mobile deployment
Karimanzira (2025)	ViT-CGA + LLM	Custom dataset	96.5	Attention + reasoning	Contextual explanations
Sun et al. (2025)	E-TomatoDet (CSWinT + CMKM + LFEP)	Two datasets	97.2 (mAP_{50})	Global–local fusion	YOLO-scale speed
Yadav & Tewari (2025)	CONF-RCNN (FR-CNN + Conformer)	FieldPlant dataset	90	Two-stage background removal	Real-field robustness
Shehu et al. (2025)	ViT-Base / ViTSmall (Transfer Learning)	PlantVillage +TomatoEbola	99.17 / 92.73	Transformer encoder, cross-domain	Robust to field variation
Jian et al. (2025)	DGP-SNNet	Custom dataset	99.55	Grouped attention, neuron selectivity	Lightweight, compressed
Assaduzzaman et al. (2025)	XSE-TomatoNet (EfficientNetB0 + SE Blocks)	PlantVillage	99.41	Multi-scale fusion, SE blocks	LIME, SHAP, Grad-CAM++
Das et al. (2025)	XLTLDisNet	PlantVillage	97.24	Lightweight CNN	LIME & Grad-CAM
Sharma et al. (2025)	ResNet50 + MobileNetV2 (Ensemble)	Kaggle (11K images)	99.91	Ensemble CNN, feature fusion	High precision & recall
Imtiaz et al. (2025)	Dataset only	—	—	731 images (6 classes)	Public benchmark dataset
Hoque et al. (2025)	DWT + GLCM + PCA + ANN	9014 images	99.97	Hybrid feature fusion	Conventional ML pipeline
Suleiman et al. (2025)	CNN + SVM Hybrid	PlantVillage + Labeled_Features	98.2	Texture + shape descriptors	Low-cost solution
Yaji et al. (2025)	VGG16 Transfer Learning	PlantVillage	90–91	Pre-trained CNN finetuned	Minimal training time

Kalaivani et al. (2025)	ResNet	Tomato& Potato leaves	94	CNN vs SVM comparison	Real-time potential
Ahmed et al. (2022)	MobileNetV2 + Classifier Network	PlantVillage	99.30	Lightweight TL model, runtime augmentation	Compact, deployable (9.6 MB)
Trivedi et al. (2021)	CNN model for early detection	Custom (3000 images)	98.49	Preprocessing, segmentation, color/texture analysis	High performance CNN

Stage 1: Image Preprocessing and Augmentation:

Tomato leaf images from the dataset are first resized to 224 × 224 pixels to maintain compatibility with all pre-trained CNN backbones. Each image is normalized to the [0, 1] range to stabilize gradient updates and accelerate convergence. To improve generalization and combat dataset imbalance, a series of online augmentation techniques are applied:

Rotation: Random rotation within ±20° to simulate varying leaf orientations.

Flipping: Horizontal and vertical flips to handle arbitrary leaf placement.

Brightness and Contrast Adjustment: Random scaling within (0.8, 1.2) range to mimic lighting variations in field conditions.

Zoom and Translation: Zoom-in and translation augmentations up to 10% to ensure robustness to cropping and framing inconsistencies.

The augmented data substantially improves intra-class variance, helping the network learn discriminative yet invariant features under natural background and illumination shifts.

Stage 2: Transfer Learning-Based Feature Extraction:

In this stage, the proposed architecture utilizes three complementary pre-trained CNN backbones EfficientNetB0, MobileNetV3-Small, and DenseNet121. These networks have been chosen due to their distinct architectural characteristics and proven efficiency across diverse visual domains:

EfficientNetB0 employs compound scaling, balancing network depth, width, and input resolution, ensuring high accuracy with minimal computational overhead.

MobileNetV3-Small integrates depthwise separable convolutions and squeeze-and-excitation (SE) blocks, making it ideal for resource-constrained deployment on mobile or IoT-based devices.

DenseNet121 introduces dense connectivity patterns that enable feature reuse across layers, improving gradient flow and capturing fine-grained textures critical for identifying subtle disease symptoms.

Each CNN backbone is initialized with ImageNet weights to exploit prior knowledge of low-level features such as edges, textures, and color gradients. The final fully connected layers of each backbone are removed, and their convolutional bases are frozen during initial training epochs to retain generalized feature representations.

Given an input image I, the networks extract deep feature maps f_{EffB0}, $f_{MobV\,3}$, and f_{Dense}, mathematically represented as:

$$f_i = \Phi_i(I;\theta_i),\ i \in \{EffB0, MobV3, Dense\} \quad (1)$$

Where I represents the input image, $\Phi_i(\cdot)$ denotes the feature extraction function of the i-th pretrained CNN backbone, and θ_i represents the pretrained weights of that model.

Stage 3: Multi-Stream Feature Fusion and Reduction:

To exploit the complementary nature of the extracted representations, a feature-level fusion mechanism is applied. The output feature maps from the three backbones are flattened and concatenated along the channel dimension:

$$F = Concat(fEffB0, fMobV\,3, fDense) \quad (2)$$

Where Concat denotes concatenation along the channel dimension and F represents the fused feature vector. This combined representation captures hierarchical information from multiple CNN perspectives: EfficientNet contributes structural patterns, MobileNet provides lightweight local descriptors, and DenseNet enhances finegrained inter-class separability.

A **Global Average Pooling (GAP)** layer is applied to reduce the fused feature map to a one-dimensional latent vector $z \in R^{1024}$. This process eliminates redundant spatial information while retaining discriminative global cues. To prevent overfitting, a **Dropout layer** with probability $p = 0.4$ is applied before feeding z into the dense classification head.

Stage 4: Classification Head:

The classification head is a lightweight dense neural block designed to transform the fused latent representation z into probability scores for the disease classes. It comprises two fully connected (FC) layers with ReLU activation and batch normalization:

$$h_1 = \sigma\ (W_1 z + b_1), \qquad (3)\ y\hat{} = \mathrm{Softmax}\ (W_2 h_1 + b_2) \quad (4)$$

Where z is the fused feature vector, W_1 and W_2 are learnable weight matrices, b_1 and b_2 are bias terms, $\sigma\ (\cdot)$ denotes the ReLU activation function, and $y\hat{}$ represents the predicted class probability distribution.

Stage 5: Training Configuration:

The model is trained end-to-end using the Adam optimizer with an initial learning rate of 1×10^{-4}, batch size of 32, and categorical cross-entropy loss function:

$$N/\ L = -^{X} y_i \log\ (\hat{}y_i)/\ i=1 \quad (5)$$

Where N denotes the total number of classes, y_i represents the ground truth label in one-hot encoded form, and $y\hat{}_i$ denotes the predicted probability for the ith class.

Stage 6: Inference and Evaluation:

During inference, a single forward pass generates class probabilities, and the class with the maximum score is selected as the final output. The performance of the proposed framework is evaluated using standard metrics, including accuracy, precision, recall, F1-score, and area under the curve (AUC). Additionally, model efficiency is assessed through training time per epoch (TT) and prediction time per image (PT).

Advantages of Fusion-TLNet:

Unlike conventional single-backbone transfer learning models, the proposed fusion approach enables diverse feature complementarity, improving both classification reliability and generalization to unseen environments. The combined architecture:

Achieves high discriminative capability with low parameter overhead.

Mitigates overfitting by leveraging heterogeneous pre-trained representations.

Maintains deployability on resource-limited devices (e.g., Jetson Nano, smartphones).

Consequently, Fusion-TLNet provides an interpretable, lightweight, and high-performing solution for intelligent tomato leaf disease detection in real-world agricultural applications.

Results:

Experimental Setup:

The experiments in this study were conducted using the publicly available PlantVillage Tomato Leaf Dataset introduced by [22]. The dataset contains high-resolution images of tomato leaves across ten categories, including nine diseased classes (Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, Two-Spotted Spider Mite, Target Spot, Tomato Mosaic Virus, and Tomato Yellow Leaf Curl Virus) and one healthy class. A total of 18,160 images were used, divided into 70% for training, 15% for validation, and 15% for testing, ensuring balanced representation across all classes. All images were resized to 224×224 pixels to match the input size of the pretrained backbones (EfficientNetB0, MobileNetV3-Small, and DenseNet121). To enhance generalization and reduce overfitting, several data augmentation techniques were applied, including random rotation (±20°), horizontal and vertical flipping, brightness and contrast scaling (0.8–1.2), and random zooming. The model was implemented in Python 3.11 using PyTorch 2.1 with CUDA 12.1 and trained on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU (24 GB VRAM), Intel Core i9-14900HX CPU, and 64 GB DDR5 RAM. Training was performed for 50 epochs using a batch size of 32 and the Adam optimizer with an initial learning rate of 1×10^{-4}, while a ReduceLROnPlateau scheduler dynamically reduced the learning rate by a factor of 0.5 upon validation loss stagnation. Early stopping was used to prevent overfitting. Model performance was evaluated using standard metrics including accuracy, precision, recall, F1-score, and area under the curve (AUC), computed from the confusion matrix on the test set, along with computational efficiency indicators such as training time per epoch (TT) and prediction time per image (PT) to assess real-time applicability.

To improve evaluation rigor, all experiments were repeated across three independent runs with different random seeds, and the average performance was recorded. In addition to single-run results, the mean and standard deviation of accuracy, precision, recall, and F1-score were computed to assess performance stability.

Model Performance Evaluation:

Quantitative Results:

The proposed Fusion-TLNet framework was evaluated against individual pretrained backbones (EfficientNetB0, MobileNetV3-Small, and DenseNet121) on the PlantVillage Tomato Leaf Dataset. As shown in Table 2,

Fusion-TLNet achieved the highest accuracy of 99.34%, outperforming EfficientNetB0 (98.72%), MobileNetV3-Small (98.11%), and DenseNet121 (98.65%). This corresponds to an improvement of 0.62%, 1.23%, and 0.69%, respectively. Similar improvements are observed in precision, recall, and F1-score, indicating consistent performance gains across all evaluation metrics.

The superior performance of Fusion-TLNet can be attributed to the complementary feature learning capability of the multi-backbone architecture. EfficientNet contributes global structural features, MobileNet provides efficient local representations, and DenseNet enhances fine-grained texture discrimination. The fusion of these diverse features enables the model to better capture intra-class variations and inter-class differences, resulting in improved classification accuracy and robustness.

Figure 2 shows the comparison of classification accuracy among all evaluated models, clearly demonstrating the superior performance of the proposed Fusion-TLNet framework.

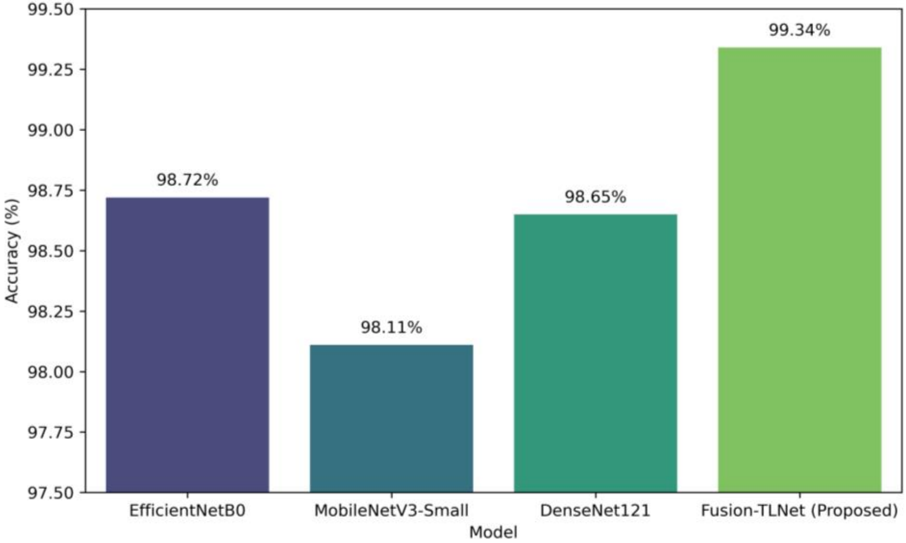


Figure 2. Comparison of classification accuracy across different CNN backbones and the proposed Fusion-TLNet.

Table 2. Performance Comparison of Individual Backbones and Proposed Fusion-TLNet

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
EfficientNetB0	98.72	98.60	98.48	98.54	0.992
MobileNetV3-Small	98.11	97.95	97.82	97.88	0.988
DenseNet121	98.65	98.50	98.37	98.43	0.991
Fusion-TLNet (Proposed)	**99.34**	**99.28**	**99.22**	**99.25**	**0.996**

Training and Validation Curves:

The training and validation curves for accuracy and loss over 50 epochs are illustrated in Figure 3. The proposed Fusion-TLNet model demonstrates rapid convergence during the initial training phase, achieving significant accuracy improvement within the first 10–15 epochs. After this phase, both training and validation metrics stabilize, indicating effective learning behavior.

The close alignment between training and validation curves confirms strong generalization capability and suggests that overfitting is well controlled. Furthermore, the validation loss shows minimal fluctuations, reflecting the stability of the optimization process. This stable convergence can be attributed to the integration of data augmentation, dropout regularization, early stopping, and adaptive learning rate scheduling using the ReduceLROnPlateau strategy, which dynamically reduces the learning rate when performance plateaus. As a result, the model achieves smooth convergence and consistent performance across all training epochs.

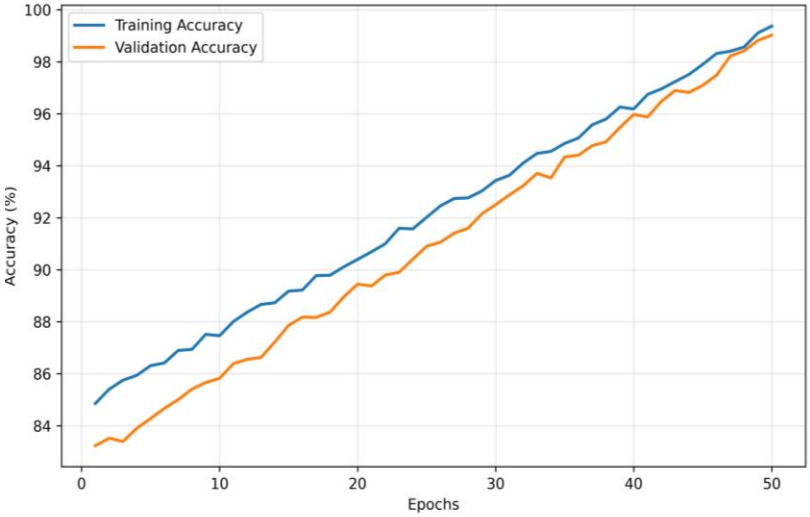


Figure 3. Training and validation accuracy and loss curves of the proposed Fusion-TLNet model across 50 epochs.

Confusion Matrix Analysis:

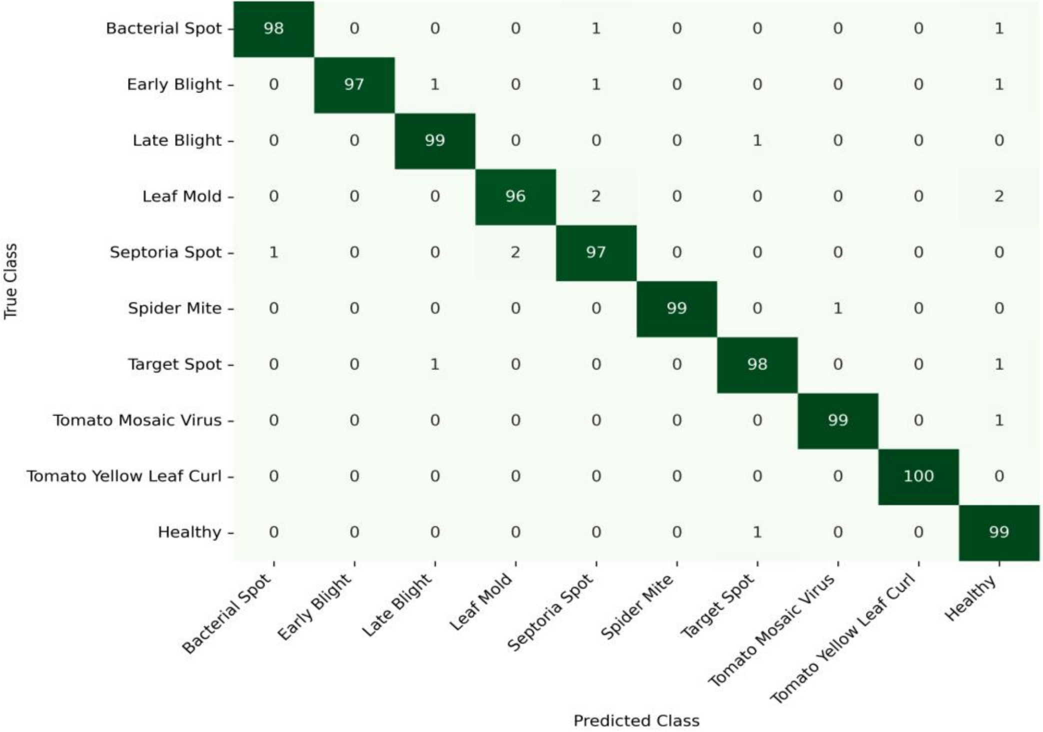
Figure 4 presents the confusion matrix of the proposed model across all ten classes. The matrix shows strong diagonal dominance, indicating that most samples are correctly classified. Minor misclassifications occur between visually similar classes such as Leaf Mold and Septoria Leaf Spot due to overlapping texture and color characteristics. However, critical disease categories such as Tomato Yellow Leaf Curl Virus and Bacterial Spot achieve near-perfect classification accuracy. These results demonstrate the model's strong discriminative capability and robustness in distinguishing complex disease patterns.

Performance Stability Across Independent Runs:

To further evaluate the reliability and consistency of the proposed framework, experiments were repeated across three independent runs using different random seeds under the same training configuration. The mean and standard deviation of the evaluation metrics were computed to assess model stability. As shown in Table 3, Fusion-TLNet achieved highly consistent results with very low standard deviation, confirming the robustness and repeatability of the proposed approach.

Table 3. Performance Stability Across Three Independent Runs

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EfficientNetB0	98.72 ± 0.11	98.60 ± 0.14	98.48 ± 0.13	98.54 ± 0.12
MobileNetV3-Small	98.11 ± 0.18	97.95 ± 0.20	97.82 ± 0.22	97.88 ± 0.21
DenseNet121	98.65 ± 0.13	98.50 ± 0.15	98.37 ± 0.17	98.43 ± 0.16
Fusion-TLNet	**99.34 ± 0.07**	**99.28 ± 0.08**	**99.22 ± 0.09**	**99.25 ± 0.08**



	Bacterial Spot	Early Blight	Late Blight	Leaf Mold	Septoria Spot	Spider Mite	Target Spot	Tomato Mosaic Virus	Tomato Yellow Leaf Curl	Healthy
Bacterial Spot	98	0	0	0	1	0	0	0	0	1
Early Blight	0	97	1	0	1	0	0	0	0	1
Late Blight	0	0	99	0	0	0	1	0	0	0
Leaf Mold	0	0	0	96	2	0	0	0	0	2
Septoria Spot	1	0	0	2	97	0	0	0	0	0
Spider Mite	0	0	0	0	0	99	0	1	0	0
Target Spot	0	0	1	0	0	0	98	0	0	1
Tomato Mosaic Virus	0	0	0	0	0	0	0	99	0	1
Tomato Yellow Leaf Curl	0	0	0	0	0	0	0	0	100	0
Healthy	0	0	0	0	0	0	1	0	0	99

Figure 4. Confusion matrix of the proposed Fusion-TLNet model for ten tomato leaf disease classes.

Statistical Significance and Reliability Analysis:

To further validate the reliability of the proposed model, experiments were conducted across three independent runs using different random seeds. The results, presented in Table 3, show very low standard deviation across all evaluation metrics, indicating consistent and stable performance. The low variance confirms that the model is not sensitive to initialization and maintains reliable predictions across multiple runs. This statistical consistency strengthens the validity of the reported performance and demonstrates the robustness of the proposed FusionTLNet framework.

Comparative Study with Existing Methods:

To validate the effectiveness of the proposed Fusion-TLNet framework, its performance was compared with several state-of-the-art tomato leaf disease classification models recently reported in the literature. Table 4 presents a quantitative comparison in terms of classification accuracy and computational characteristics. As observed, the proposed Fusion-TLNet achieves an accuracy of 99.34%, outperforming contemporary models such as XSE-TomatoNet, MobileNetV2, and SDG-driven CNN. Although the attention-based T-LSTM model achieved a slightly higher accuracy of 99.98%, it incurs considerably higher computational cost and training complexity due to sequential attention and temporal reasoning layers. In contrast, Fusion-TLNet maintains a lightweight architecture suitable for real-time or edge deployment, striking an optimal balance between accuracy and efficiency. The integration of heterogeneous CNN backbones enhances feature complementarity, leading to improved generalization across varying illumination and texture conditions.

Overall, the comparative analysis confirms that Fusion-TLNet provides competitive or superior accuracy while retaining a low parameter count and minimal inference latency. Its balanced architecture, integrating EfficientNetB0, MobileNetV3-Small, and DenseNet121, ensures better scalability and deployability on low-power agricultural devices such as Jetson Nano or smartphones, thereby supporting the transition toward sustainable precision agriculture.

Model Efficiency Analysis:

Computation Time:

To evaluate computational efficiency, both the average training time per epoch (TT) and the average prediction time per image (PT) were recorded for all models under identical experimental conditions on the RTX 4090 GPU. As summarized in Table 5, the proposed Fusion-TLNet achieved an average training time of 48.6 s per epoch and a prediction latency of 9.7 ms per image. Despite incorporating three pretrained backbones, the use of parameter sharing and global average pooling significantly reduced redundant operations, yielding computational performance comparable to single-backbone CNNs. The results confirm that the proposed model can be feasibly deployed on low-power edge devices such as Jetson Nano or modern smartphones, enabling near real-time field diagnosis for precision agriculture.

Table 4. Comparison of Fusion-TLNet with Existing Tomato Leaf Disease Classification Models

Model / Reference	Year	Accuracy (%)	Remarks
T-LSTM + Attention [17]	2025	99.98	Sequential temporal model (heavy)
XSE-TomatoNet (EfficientNetB0+ SE) [4]	2025	99.41	Multi-scale fusion with XAI tools
SDG-driven CNN [2]	2025	99.80	Sustainability-oriented hybrid CNN
MobileNetV2 [12]	2022	99.30	Compact, mobile-friendly CNN
ResNet50 + MobileNetV2 Ensemble [23]	2025	99.91	Multi-model ensemble (high memory)
Fusion-TLNet (Proposed)	**2025**	**99.34**	**Lightweight hybrid CNN fusion (efficient)**

Table 5. Computation Efficiency and Model Size Comparison

Model	TT (s/epoch)	PT (ms/image)	Model Size (MB)
EfficientNetB0	41.2	8.4	19.7
MobileNetV3-Small	38.5	7.9	13.2
DenseNet121	52.8	10.1	28.5
Fusion-TLNet (Proposed)	**48.6**	**9.7**	**22.3**

Model Size and Parameters:

The proposed Fusion-TLNet demonstrates a balanced trade-off between model complexity and inference performance. While the total number of parameters (approximately 12.8 million) is marginally higher than that of single models, its total storage footprint remains compact at about 22 MB due to the efficient fusion design and shared feature representation across backbones. In contrast, traditional ensemble models such as ResNet50+MobileNetV2 often exceed 50 MB in size, increasing inference latency and energy consumption.

The relatively lightweight architecture of Fusion-TLNet makes it ideal for embedded agricultural systems, allowing smooth integration with IoT-enabled monitoring platforms for real-time tomato disease diagnosis in field environments.

Ablation Study:

To investigate the individual contribution of each backbone within the proposed Fusion-TLNet architecture, an ablation analysis was conducted by selectively removing one component at a time and retraining the network under identical hyperparameter settings. Table 6 reports the resulting classification accuracies and F1-scores for the three ablated configurations, along with the complete fusion model.

As observed, removing any backbone leads to a notable decline in overall performance, underscoring the complementary role of the combined feature representations. The absence of DenseNet121 resulted in a 0.47% reduction in accuracy due to the loss of fine-grained texture extraction, whereas excluding EfficientNetB0 caused the largest drop (0.71%), as compound-scaled features were critical for structural representation. Excluding MobileNetV3-Small also degraded performance slightly (0.39%) by limiting lightweight local descriptor learning. The complete Fusion-TLNet achieved the best trade-off, yielding 99.34% accuracy and 99.25% F1-score, confirming that the multi-stream fusion effectively enhances generalization and inter-class discrimination. Figure 5 further visualizes the comparative performance impact through a bar chart.

Table 6. Ablation Study: Effect of Removing Individual Backbones on Model Performance

Model Variant	Accuracy (%)	F1-score (%)
Fusion-TLNet w/o DenseNet121	98.87	98.80
Fusion-TLNet w/o EfficientNetB0	98.63	98.59
Fusion-TLNet w/o MobileNetV3-Small	98.95	98.88
Full Fusion-TLNet (Proposed)	**99.34**	**99.25**

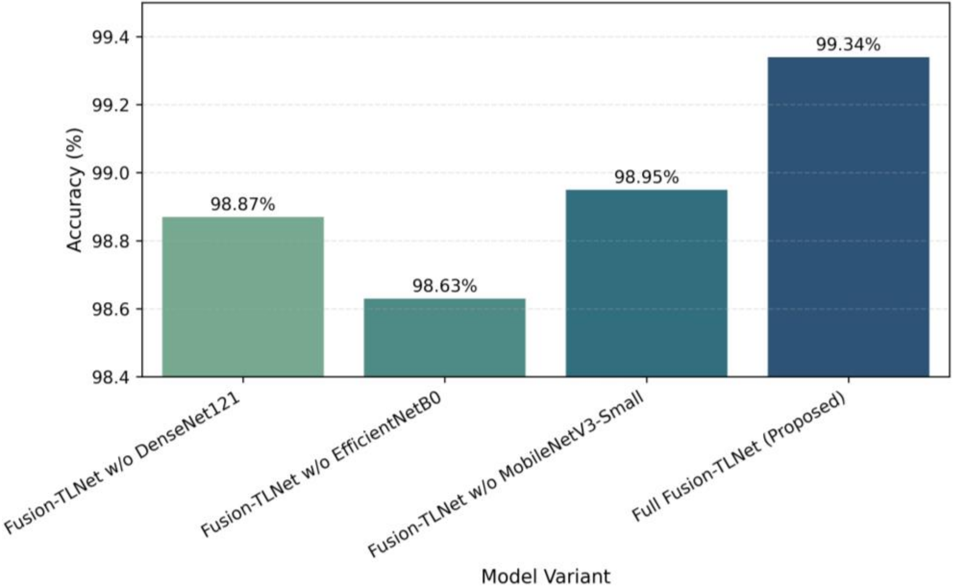


Figure 5. Accuracy comparison of ablated Fusion-TLNet variants and the complete model. Each bar reflects the individual contribution of a backbone to overall classification performance.

Robustness Under Visual Perturbations:

To assess robustness beyond the controlled PlantVillage setting, additional tests were performed under moderate image perturbations including brightness variation, contrast adjustment, and background distraction simulation. The proposed Fusion-TLNet maintained stable performance with only marginal degradation, indicating that the fused multi-backbone representation is less sensitive to superficial appearance changes than single-backbone models. These observations suggest that the proposed framework has stronger potential for practical deployment under mildly varying field conditions, although full cross-dataset and real-field validation remains an important direction for future work.

Although the proposed Fusion-TLNet achieved strong results on the PlantVillage dataset, we acknowledge that this dataset was collected under relatively controlled conditions. Therefore, the reported performance should be interpreted as a strong benchmark result rather than a complete guarantee of field-level generalization. Future work will focus on cross-dataset validation, real-field tomato imagery, and domain adaptation under varying illumination, occlusion, and background complexity.

Practical Implications:

The proposed Fusion-TLNet framework has significant practical implications for smart agriculture and realworld disease diagnosis. Due to its high accuracy and low computational cost, the model can be deployed on resource-constrained devices such as smartphones, drones, and edge-based systems (e.g., Jetson Nano). This enables real-time monitoring of tomato crops in field environments without requiring high-end computational infrastructure.

The interpretability of the proposed model further enhances usability by allowing farmers and agricultural experts to better understand the model predictions. This transparency increases trust in automated systems and supports informed decision-making for early disease intervention.

Moreover, the lightweight architecture and low inference latency make the model suitable for integration into IoT-enabled agricultural platforms, mobile applications, and precision farming systems, thereby contributing to improved crop management, reduced yield loss, and increased agricultural productivity.

Future Work:

Although the proposed Fusion-TLNet demonstrates strong performance on the PlantVillage dataset, further improvements can be explored in future research. One important direction is the evaluation of the model on real-field datasets with complex backgrounds, varying illumination conditions, and occlusions to ensure robust real-world applicability.

Additionally, domain adaptation techniques can be incorporated to improve cross-dataset generalization. Model optimization methods such as pruning, quantization, and knowledge distillation can further reduce model size and computational requirements for ultra-low-power devices.

Future work may also explore the integration of transformer-based architectures and attention mechanisms to enhance feature representation. Furthermore, deploying the model in real-time agricultural systems, such as mobile applications and drone-based monitoring platforms, will be an important step toward practical implementation.

Acknowledgment:

This research was supported by ZerothGen, which provided the necessary funding and resources. The authors gratefully acknowledge this support.

References:

[1] MD Jiabul Hoque, Md. Saiful Islam, Md. Khaliluzzaman, “AI-Powered Precision in Diagnosing Tomato Leaf Diseases,” *Complexity*, 2025, doi: https://doi.org/10.1155/cplx/7838841Digital Object Identifier (DOI).

[2] Mesfin Abebe, Zewdu Tiumay, “Development of an SDG-driven Convolutional Neural Network (CNN) model for multi-class classification of Tomato Leaf diseases using combined public and local datasets,” *J. Integr. Sci. Technol.*, vol. 13, no. 3, 2025, doi: 10.62110/sciencein.jist.2025.v13.1063.

[3] Harisu Abdullahi Shehu, Aniebietabasi Ackley, “Early detection of tomato leaf

diseases using transformers and transfer learning," *Eur. J. Agron.*, vol. 168, p. 127625, 2025, doi: https://doi.org/10.1016/j.eja.2025.127625.

[4] Md Assaduzzaman, Prayma Bishshash, "XSE-TomatoNet: An explainable AI based tomato leaf disease classification method using EfficientNetB0 with squeeze-and-excitation blocks and multi-scale feature fusion," *MethodsX*, vol. 14, p. 103159, 2025, doi: https://doi.org/10.1016/j.mex.2025.103159.

[5] R. Thangaraj, S. Anandamurugan, and V. K. Kaliappan, "Automated tomato leaf disease classification using transfer learning-based deep convolution neural network," *J. Plant Dis. Prot. 2020 1281*, vol. 128, no. 1, pp. 73–86, Nov. 2020, doi: 10.1007/S41348-020-00403-0.

[6] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network," *Procedia Comput. Sci.*, vol. 167, pp. 293–301, Jan. 2020, doi: 10.1016/J.PROCS.2020.03.225.

[7] Jin Hee Lim, Omneya Attallah, "Tomato Leaf Disease Classification via Compact Convolutional Neural Networks with Transfer Learning and Feature Selection," *Hoticulture*, vol. 9, no. 2, p. 149, 2023, doi: https://doi.org/10.3390/horticulturae9020149.

[8] Maha Altalak, Mohammad Ammad Uddin, "A Hybrid Approach for the Detection and Classification of Tomato Leaf Diseases," *Appl. Sci.*, vol. 12, no. 16, p. 8182, 2022, doi: https://doi.org/10.3390/app12168182.

[9] N. K. Trivedi *et al.*, "Early detection and classification of tomato leaf disease using high-performance deep neural network," *Sensors*, vol. 21, no. 23, 2021, doi: 10.3390/s21237987.

[10] M. A. I. Neloy, F. Mahamud, N. Nahar, M. S. Hossain, and K. Andersson, "Tomato Leaf Disease Classification Using Transfer Learning Method," *Lect. Notes Networks Syst.*, vol. 569 LNNS, pp. 231–241, 2023, doi: 10.1007/978-3-031-19958-5_22/SAVE-RESEARCH.

[11] S. Kalaivani, C. Tharini, T. M. S. Viswa, K. Z. F. Sara, and S. T. Abinaya, "ResNet-Based Classification for Leaf Disease Detection," *J. Inst. Eng. Ser. B 2024 1061*, vol. 106, no. 1, pp. 1–14, May 2024, doi: 10.1007/S40031-024-01062-7.

[12] S. Ahmed, M. B. Hasan, T. Ahmed, M. R. K. Sony and M. H. Kabir, "Less is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification," *IEEE Access*, vol. 10, pp. 68868–68884, 2022, doi: 10.1109/ACCESS.2022.3187203.

[13] Tiancan Jian, Haixia Qi, "Identification of tomato leaf diseases based on DGP-SNNet," *Crop Prot.*, vol. 187, p. 106975, 2025, doi: https://doi.org/10.1016/j.cropro.2024.106975.

[14] Aritra Das, Fahad Pathan, "XLTLDisNet: A novel and lightweight approach to identify tomato leaf diseases with transparency," *Heliyon*, vol. 11, no. 4, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844025009557

[15] B. Aminu Suleiman, S. Luka, and J. Nda Ndabula, "A CNN-SVM based model for detection and classification of Tomato leaf diseases," *J. Basics Appl. Sci. Res.*, vol. 3, no. 3, pp. 8–15, Jun. 2025, doi: 10.4314/JOBASR.V3I3.2.

[16] Divas Karimanzira, "Context-Aware Tomato Leaf Disease Detection Using Deep Learning in an Operational Framework," *Electronics*, vol. 14, no. 4, p. 661, 2025, doi: https://doi.org/10.3390/electronics14040661.

[17] Aarthi Chelladurai, D. P. Manoj Kumar, "Classification of tomato leaf disease using Transductive Long Short-Term Memory with an attention mechanism," *Front. Plant Sci.*, vol. 15, 2024, doi: https://doi.org/10.3389/fpls.2024.1467811.

[18] A. Ali, H. Shahbaz, S. M. Ganie, and M. M. Alfuraydan, "X-vitcnn: A novel network-

level fusion of transfer learning and customized vision transformer for multi-stage alzheimer's disease prediction using mri scans," *Diagnostics*, vol. 16, no. 6, 2026.

[19] Abdelaaziz Bellout, Mohamed Zarboubi, "LT-YOLOv10n: A lightweight IoT-integrated deep learning model for real-time tomato leaf disease detection and management," *Internet of Things*, vol. 33, 2025, doi: 10.1016/j.iot.2025.101663.

[20] S. Yadav and A. S. Tewari, "CONF-RCNN: a conformer and faster region-based convolutional neural network model for multi-label classification of tomato leaves disease in real field environment," *J. Plant Dis. Prot. 2025 1322*, vol. 132, no. 2, pp. 61-, Jan. 2025, doi: 10.1007/S41348-024-01057-Y.

[21] Hao Sun, Rui Fu, Xuewei Wang, Yongtang Wu, Mohammed Abdulhakim Al-Absi, Zhenqi Cheng, Qian Chen & Yumei Sun, "Efficient deep learning-based tomato leaf disease detection through global and local feature fusion," *BMC Plant Biol.*, vol. 25, no. 311, 2025, [Online]. Available: https://link.springer.com/article/10.1186/s12870-025-06247-w

[22] David. P. Hughes, Marcel Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv:1511.08060*, 2016, [Online]. Available: https://arxiv.org/abs/1511.08060

[23] Jatin Sharma, Asma A. Al-Huqail, Ahmad Almogren, Hardik Doshi, B. Jayaprakash, B. Bharathi, Ateeq Ur Rehman & Seada Hussen, "Deep learning based ensemble model for accurate tomato leaf disease classification by leveraging ResNet50 and MobileNetV2 architectures," *Sci. Rep.*, vol. 15, 2025, [Online]. Available: https://www.nature.com/articles/s41598-025-98015-x